

NAVAL HEALTH RESEARCH CENTER

ORGANIZATIONAL INFLUENCES ON GENDER DIFFERENCES IN STRESS AND STRAIN ABOARD U.S. NAVY SHIPS

R. R. Vickers, Jr.

J. A. Martin

Report No. 98-33

19991108 082

DTIC QUALITY INSPECTED 4

Approved for public release; distribution unlimited.



NAVAL HEALTH RESEARCH CENTER
P O BOX 85122
SAN DIEGO, CA 92186-5122
BUREAU OF MEDICINE AND SURGERY (MED-02)
2300 E ST. NW
WASHINGTON, DC 20372-5300



Organizational Influences on Gender Differences in
Stress and Strain Aboard U.S. Navy Ships

Ross R. Vickers, Jr.¹
James A. Martin²

¹Human Performance Department
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92186-5122

and

²School of Social Work
Bryn Mawr College
Bryn Mawr, PA

Report 98-33 was supported by the U.S. Navy Medical Research and Development Command through funds provided by the U.S. Army Medical Research and Materiel Command under Work Unit 6611. The views expressed are those of the authors and do not reflect the official policy or the position of the Department of the Navy, Department of the Army, Department of Defense, or the U.S. Government. Approved for public release; distribution unlimited.

The Naval Health Research Center Committee for the Protection of Human Subjects approved the use of human subjects under the provisions of SECNAVINST 3900.39b, NMRDCINST 3900.2, and NHRCINST 3900.2. These instructions specify the Common Rule of Federal Policy for the Protection of Human Subjects, 32 CFR 219, as the minimum protection standard for Department of Navy research.

SUMMARY

Problem

Previous research has shown that women assigned to shipboard duty in the U.S. Navy generally report only slightly higher psychological stress and strain than their male shipmates. However, average trends might mask substantial gender differences between selected groups of men and women.

Objective

The present study investigated whether ship assignment, Navy rating, or previous deployment experience was related to the magnitude of gender differences in stress.

Approach

Gender differences in stress and strain items in the Women Aboard Navy Ships survey were determined for each of 11 ship types and 24 Navy ratings. The differences also were determined for personnel assigned to ships either in home port or away from home port. The differences observed aboard each ship, in rating, and for each port status were compared to the difference for the whole sample. Statistical criteria were applied to determine whether the differences were significantly larger or smaller in each subgroup than in the sample as a whole.

Results

Overall, women generally reported slightly higher stress and strain than men, but the differences were small on the average. However, women were substantially more concerned about children and child care than men. Deployment experience did not affect the size of the gender differences, but ship type and rating did. In each of these cases, however, the effects were ambiguous. Some ratings and ship types amplified the full sample trend toward higher stress and strain among women, but other ratings and ship types reversed this trend. On the whole, there were more reversals than amplifications.

Conclusions

Shipboard gender differences in stress and strain are small. Stresses arising from work-family conflict were the notable exception to this general trend. This pattern of findings was very similar to results obtained in studies of civilians. Ship type and rating influenced the size of gender differences, but these moderator effects were scattered. The differences that were identified should be replicated before they are taken too seriously because there was no theory to predict which ship types or ratings would yield significant moderator effects. The observed moderator effects could be the result of chance given that a large number of groups were evaluated. In summary, shipboard gender differences in stress and strain in the U.S. Navy are small and similar to differences found in the general civilian population.

INTRODUCTION

Background

Shipboard living and working conditions generate significant stress for U.S. Navy personnel (Martin, Acevedo, & Garland, in review). Understanding the pattern of stress and the processes that produce stress is necessary in order to reduce stress. Earlier work in the Women Aboard Navy Ships (WANS) study indicates that stress and strain differences between male and female crew members were slight on the average (Martin et al., in review; Vickers & Martin, in review). However, ship type and Navy occupation are known to influence organizational climate and habitability variables that are similar to the stresses investigated in the WANS study (Rubin, Gunderson, & Arthur, 1971; Pugh, Gunderson, & Dean, 1975). The object of this study was to determine whether these organizational factors affected the magnitude of gender differences in stress.

Previous Research on Gender and Stress

The initial finding that shipboard female-male stress differences were minor may be consistent with prior research. The uncertainty in this statement arises because previous reviews have reached conflicting conclusions. Martocchio and O'Leary (1989) reported a very small average correlation between gender and stress ($r = -.02$; $n = 7,884$). In contrast, Jick and Mitz (1985) concluded that women reported more psychological stress symptoms than men, but men were more susceptible to physical problems that might be caused by stress (e.g., coronary disease).

The apparent conflict between the conclusion reached by Jick and Mitz and that reached by Martocchio and O'Leary (1989) may be the effect of methodological differences in the two reviews. Jick and Mitz (1985) based their conclusion on the fact that a large proportion of the studies they reviewed demonstrated statistically significant differences between women and men. However, 11 of 19 of the studies in their review had samples of 1,000 or more participants. Given these sample sizes, even small differences would be statistically significant. Jick and Mitz (1985) also concentrated on symptomatic measures, including affective states and psychosomatic symptoms. The Martocchio and O'Leary (1989) review included more measures that would be considered situational antecedents of these symptomatic responses. Perhaps the antecedent conditions are less strongly differentiated between men and women than are the responses to those conditions.

Recent evidence suggests that both factors mentioned in the preceding paragraph are relevant when considering gender differences in stress. Women generally have reported slightly higher stress than men on the average, but the differences have been quite small on the average (Appendix A). When each difference is expressed as a point biserial correlation (r_{pb}), 66% of the correlations are between $r_{pb} = -.10$ and $r_{pb} = .10$. The distribution is slightly skewed toward positive values so the median value is $r_{pb} \approx .03$. Thus, both Martocchio and O'Leary's (1989) review and recent data point to very small differences between women and men. Just how small the typical difference is can be seen by noting that Cohen (1988) sets $|r_{pb}| = .10$ as the lower bound for an effect that is large enough to be of practical or theoretical importance (cf., Cohen,

1988). Neither Martocchio and O'Leary's (1989) review nor the present review suggest that gender differences in stress will approach even this minimum criterion.

Detailed consideration of the evidence supports the suggested substantive basis for the conflict between the Jick and Mitz (1985) and Martocchio and O'Leary (1989) conclusions. It has long been recognized that "stress" is a very broad construct encompassing a diverse range of stress indicators (Hinkle, 1973). Some theorists divide the general domain into "stress" and "strain." When this is done, "stress" is broadly defined as events or environmental conditions that require effort or adaptation on the part of the individual. Holmes and Rahe's (1967) life events scale is the classic event approach to quantifying stress. Caplan, Cobb, French, Harrison, and Pinneau (1975) provide an excellent example of defining stress in terms of environmental conditions such as role ambiguity, overload, and responsibility on the job. "Strain" then is defined as emotional, behavioral, or biological states of the individual that are reactions to the events and conditions comprising stress.

Given a distinction between stress and strain, Jick and Mitz's (1985) conclusions can be regarded as indicating that women report more psychological strain than men. Recent studies support this assertion. Women generally report higher depression and anxiety than do men (e.g., Brown & Gary, 1987; Conger, Lorenz, Elder, Simons, & Ge, 1993; Hendrix et al., 1994; Lester, Posner, & Leitner, 1986; Sowa & Lustman, 1984; Thoits, 1987; Turner & Avison, 1989; Salgado de Snyder et al., 1990; Greenglass & Burke, 1988; Waelde, Silvern, & Hodges, 1994), although some studies yield smaller, statistically nonsignificant differences (Caldwell, Pearson, & Chin, 1987; Cushway, 1992; Richman & Flaherty, 1990; Wu, 1996; Zuckerman, 1989; McCreary et al., 1996; Heppner, Cook, Strozier, & Heppner, 1991). Taken as a set, however, the typical result is a small, but potentially important, difference in which women report more strain than men (median $r_{pb} = .13$ for depression scales, cf. Appendix A). This observation is consistent with Jick and Mitz's (1985) conclusion that women experience greater psychological stress than men. The implication is that Martocchio and O'Leary's (1989) conclusion may represent an over generalization about gender differences in stress.

Appropriate conclusions about gender differences in stress also may be reached only if stress indicators assessing environmental conditions are subdivided. Some stressful events and conditions may show consistent gender differences. Work-family conflict provides a specific example of this need. Women commonly experience more conflict between work and family demands than do men (Bedeian, Burke, & Moffett, 1988; Beutell & O'Hare, 1987; Sekaran, 1986; Salgado de Snyder, Cervantes, & Padilla, 1990; Greenglass & Burke, 1988). Exceptions to this trend have occurred (Hendrix, Spencer, & Gibson, 1994; Swanson, Power, & Simpson, 1996), but the bulk of the evidence supports the position that women experience more stress in this area than men. Most differences have been in the $.10 < r_{pb} < .30$ range (median $r_{pb} = .16$; cf., Appendix A).

Finally, prior research provides little information about the direction or size of gender differences for some stresses. The potential for using prior research findings to draw inferences about the U.S. Navy shipboard population also is hampered by the fact that some potentially important stressful conditions have not been studied in other settings.

During deployment, a ship is a total institution, i.e., all aspects of each sailor's life take place within the confines of the ship (Goffman, 1961). As a result, issues such as crowding, privacy, limited recreational opportunities, and other factors are potential sources of stress that may be uncommon in other settings. Crowding has been studied in dormitory settings (e.g., Mandel, Baron & Fisher, 1980; Walden, Nelson & Smith, 1981). Gender differences in crowding apparently are small on the average, but there may be important interactions between gender and specific aspects of the environment. While the finding that average gender differences in crowding are negligible is consistent with the general trend in research on gender reactions to environmental conditions, the relevance of such studies to the present findings is questionable. There are substantial differences between dormitory life and shipboard life, including degree of crowding and the availability of other opportunities for privacy. If gender differences in what might be called habitability stresses are subject to subtle interactions, as the data suggest, generalizing from one setting to another markedly dissimilar setting is risky. Thus, an effective understanding of gender differences in stress may require consideration of specific situations for at least some types of stress. At present, there are too few data available to identify the relevant factors that influence the size of gender differences in stress.

Possible Moderators of Stress Effects

The preceding review of prior evidence from studies of the civilian population and the previous evidence from the WANS study suggests two conclusions: Gender differences in stress are virtually nonexistent, except for work-family conflict. Gender differences in strain may be reliable, but they are weak.

The two general conclusions may have limited relevance to the U.S. Navy shipboard population. The observations are based on studies that aggregated data from many different types of people. The aggregate data will be misleading if there are subgroups of individuals within a larger population for which gender differences in stress and strain are substantial. The fact that men and women in the WANS study differed little in overall stress and strain (Martin et al., in review) does not rule out the possibility that there are identifiable subgroups within the large population for which those differences are large. In fact, the overall trend may be the result of having one or more subsets within the population who experience particularly high stress while there is no difference between women and men in the remainder of the population. Identifying specific factors affecting the size of gender differences in stress, if such factors exist, can lead to insights that can be used to reduce that stress.

This study examined three potential moderators of gender differences in stress levels experienced by personnel assigned to U.S. Navy ships:

A. Occupation. Women have had increasing opportunities to fill all types of jobs in the military in recent years. However, the proportion of women varies across occupational specialities. When the proportion of women is low, women may face a different social environment and new tasks without preexisting support or a mentoring network to help them adjust to the situation. The available evidence suggests that there are

limited differences between men and women in the same occupation, but that evidence is limited to professional and white collar jobs. The situation may differ when extended to work performed in a more industrial setting with heavier physical demands and other stressful working conditions.

B. Ship Type. Living conditions are an important source of shipboard stress (Pugh et al., 1975). Prior evidence indicates that these conditions vary as a function of ship (Pugh et al., 1975). If women are more susceptible to some sources of stress than men, pronounced experiential differences may occur only when stress-eliciting conditions are severe. The evidence that there are no strong gender differences in perceived crowding suggests that these conditions will not affect men and women differently. However, there is an obvious risk in generalizing from one specific aspect of housing in open civilian dormitory settings to a wide range of potential stress conditions aboard a U.S. Navy ship.

C. Deployment Effects. Navy personnel are exposed to some unusual stresses because the job often requires absence from one's home and family during deployments. Some stresses may occur only during deployment or may be exacerbated by deployment. For example, if one ordinarily does not live aboard ship, shipboard living conditions become relevant only during deployment. Similarly, family stresses may be increased (e.g., by making it more difficult to communicate) or decreased by deployment (e.g., by removing the individual from a difficult social relationship). Deployment may increase stress levels because the ship becomes a total institution encompassing all aspects of a person's life (Goffman, 1961). If so, deployment can be a stress multiplier because different types of stress reinforce one another (e.g., by increasing carryover from work to leisure settings). The impact of deployment on an individual's stress may vary depending on prior deployment experience. However, the direction of any such effects is uncertain. In most cases, it could be argued that past experience could act to either increase stress (e.g., because the deployment itself is the cause of stress) or to decrease it (e.g., because the person develops effective methods of coping with stress). In light of this uncertainty about the direction of effects, the most reasonable a priori expectation is provided by generalizing from the overall body of evidence on gender differences in stress. That evidence suggests that past experience will not affect the size or nature of gender differences in stress.

Hypotheses

Based on the preceding considerations, the present study tested three general hypotheses:

A. Women will report higher stress and strain than men, but the typical difference will be too small to be theoretically or practically important.

B. Work-family conflict and depression will be areas in which women will report higher stress and strain than men; these differences will be large enough to be important.

C. Occupation, ship type, and deployment status will not affect the size of gender differences in stress and strain.

The first two hypotheses are based on trends in the stress-gender literature summarized in Jick and Mitz (1985), Martocchio and O'Leary (1989), and Appendix A of this report. The third hypothesis adopts the conservative position of assuming that occupation, ship type, and deployment status have no effect until there is evidence to reject this position.

METHOD

Sample

The sample consisted of female sailors and male sailors matched to the women on work division, department, ethnicity, pay grade, rating, and age (Garland & Timberlake, 1995). These sailors completed either Form 7, 8, or 9 of the WANS questionnaire. This subset of the overall WANS study consisted of 2,822 sailors. The average age of these sailors was 26.3 (SD = 6.2) years. The typical educational level was a high school graduate (52.0%) or Graduate Equivalency Diploma (3.7%). A substantial minority (32.4%) indicated "Some college/Associate of Arts" and 5.8% had a Bachelor's degree. Sailors with graduate degrees were as common as sailors with less than a high school education (1.2% each). Finally, a few sailors had attended trade or technical school after high school (3.6%). Marital status typically was "Never married" (46.6%) or "Married" (41.4%). Most others in the sample were "Separated" (5.0%) or "Divorced" (6.8%) with a very small minority "Widowed" (0.3%).

The sample was ethnically diverse. "Non-Hispanic white" was the most common ethnic designation chosen by the respondents (53.0%), followed by "Non-Hispanic black" (29.6%). Hispanics comprised 7.4% of the total (the sum of 5.3% who designated themselves as "White, Hispanic" and 2.1% who designated themselves "Black, Hispanic"). "Asian/Pacific Islander" was chosen by 3.5% of the sample and "Native American" by 1.2%. The remaining 5.4% of the sample chose "Other race/ethnicity."

Pay grades were widely distributed as would be expected given the rank structure of the U.S. Navy and the sampling procedures. The great majority of the sample were enlisted (95.6%) rather than officer (4.4%). Few enlisted personnel were below the E-3 pay grade (E-1, 3.0%; E-2, 10.7%). E-3s (18.5%), E-4s (26.0%), and E-5s (19.4%) comprised the majority of the sample (63.9%). E-6s (12.8%) were three times as common as E-7s (4%). E-8s (0.8%) and E-9s (0.4%) each represented less than 1% of the sample.

Sample size varied for different stress and strain items, primarily because items about spouse and children applied to only some sailors. The quality of life items for spouse (n = 1,294) and children (n = 1,191) had small sample sizes compared to other quality of life items (n ≥ 2,740). Stress items for dealing with children (n = 1,392), discipline of children (n = 1,374), and child-care arrangements (n = 1,358) had small sample sizes compared to other stress items (n ≥ 2,452).

Stress Questionnaire

Stress and distress were measured by self-report items in the WANS questionnaire (Appendix B). The stress/distress content of the WANS questionnaire consisted of a 50-item self-report questionnaire. Stresses were assessed by 36 items, 33 describing specific conditions of Navy shipboard life and 3 describing effects of those conditions on the individual's personal life, job performance, and coping. Stress item responses were made on a 5-point Likert scale with options ranging from "None at all" (scored "1") to "Extreme amount" (scored "5").

An additional 14 items were treated as strains, i.e., presumed consequences of stress. Seven of the items dealt with quality of life (QoL) in the Navy. These items were answered on a 7-point scale with options ranging from "Terrible" (scored "1") to "Delighted" (scored "7") with "Mixed" (scored "4") as the midpoint of the scale. Seven depression items from the CES-D scale (Radloff, 1977) were the other strain indicators. Responses to these items employed an 8-point scale indicating the number of days during the past week the person had experienced the feelings described in the item. Response options ranged from "No days" (scored "0") to "Seven days" (scored "7").

Scores for the seven quality of life items and the single coping item were reversed prior to computing associations to gender. The reversals meant that a high score meant high stress or strain for all items.

Rating

Shipboard occupation was defined by responses to a question asking the respondent "If you are Navy enlisted, what is your rating?" Respondents were asked to enter a two- or three-letter code for their rating. This code was entered in the data set and used to define rating membership. For the analyses, rating classification was based on the first two letters of the code (e.g., SK for storekeeper, GM for Gunner's Mate). Greater detail could have been obtained by retaining some three-letter codes, but small sample sizes would have made it necessary to default to the two-letter code in most cases in any event. Using the two-letter code, therefore, provided a uniform level of classification with minimal likelihood of underutilizing the stress data.

Deployment Status

The deployment status of individuals in the study was determined from two questions in the WANS survey. One question asked "If you are currently aboard ship, what is your ship's current status?" Response options were "In home port," "At sea," "In port other than home port," "In shipyard," and "Other." The large majority of respondents indicated they were "In home port," so this variable was recoded to a two-level indicator. The two levels were home port and all other locations.

Deployment experience was evaluated by combining responses to two questions. The first question was "How many times have you deployed aboard Navy ships (30 days or more), not counting the present deployment?" If the respondent had never deployed, he/she filled in a blank marked "0" on the questionnaire. If the respondent had deployed one or more times, he/she was asked to write in the number of deployments.

Responses to the second question ranged from 1 to more than 90. These responses were recoded to yield a six-level variable with levels of 0 deployments, 1 deployment, 2 deployments, 3 or 4 deployments, 5 to 8 deployments, and 9 or more deployments.

Ship Type

U.S. Navy ships are classified into ship types based on mission functions (e.g., aircraft carriers, destroyers, oilers, ammunition ships). Ships within a type have similar designs to support their mission functions. Ships within type have similar working and living conditions, but those conditions differ between types (Pugh et al., 1975). Differences between types with respect to mission, physical configuration of the ships, and other factors that differ systematically between types could influence stress and strain. This possibility was tested by comparing the magnitude of gender differences across nine ship types: (1) Guided Missile Destroyer (DDG; 2 ships), (2) Fast Combat Support Ship (AOE; 5 ships), (3) Amphibious Dock Landing Ship (LSD; 2 ships), (4) Miscellaneous Command Ship (AGF; 3 ships), (5) Salvage Ship (ARS; 3 ships), (6) Ammunition Ship (AE; 5 ships), (7) Oiler (AO; 4 ships), (8) Destroyer Tender (AD; 3 ships), (9) Submarine Tender (AS; 7 ships).¹

Analysis Procedures

Analyses were conducted with the SPSS statistical package (SPSS, Inc., 1992) except where otherwise noted. Gender differences were expressed as r_{pb} values with item responses treated as (approximately) continuous variables and gender as a dichotomy. These correlations can be interpreted directly as effect sizes by applying Cohen's (1988) criterion² of $|r| = .10$ as the lower boundary for a correlation that would have theoretical or practical significance. An r_{pb} of this magnitude can be interpreted using Rosenthal and Rubin's (1982) binomial effect size display (BESD) that restates point biserial correlations as

$$r_{pb} \approx \frac{\text{Proportion of women with above average scores} - \text{Proportion of men with above average scores}}{1}$$

In this form, a correlation at Cohen's (1988) lower boundary for theoretical or practical significance implies that 55% of women scored above average on a stress compared to 45% of men (or vice versa).

The r_{pb} values were analyzed further to identify moderator effects:

- A. Frequency distributions for the potential moderators were examined to identify categories with sufficiently large sample sizes for further reasonably accurate estimation of correlations.
- B. The gender differences were translated into r_{pb} values for each stress and strain item in each subgroup defined by the moderator.
- C. Fisher's r -to- z transformation was applied to each correlation. A pooled within-group correlation was computed by taking the weighted average of these transformed correlations.
- D. Hays' (1963) V was computed for each correlation. This computation provided a χ^2 suitable for testing the hypothesis that the

correlation demonstrated greater than chance variation across the groups. For each item, the correlations used to compute V would be distributed as a χ^2 with $k - 1$ degrees of freedom (where k is the number of groups for the moderator) if the null hypothesis were true.

E. The item χ^2 s were summed to produce cumulative χ^2 s suitable for testing the overall effects of the moderator. Cumulative χ^2 s were computed separately for the items pertaining to quality of life, depression, and stress.

F. The differences between the correlations within a specific group and the pooled correlations were examined. This step was necessary because each cumulative χ^2 was an omnibus test involving more than one degree of freedom (Rosenthal & Rosnow, 1984, pp. 471 and 473). The examination began by computing the average difference for the set of items

$$d. = \sum d_i / a = \sum (r_{ig} - r_{ip}) / a$$

where $d.$ is the average difference, d_i is the difference for the i th item, r_{ig} is the point biserial correlation for the i th item in the g th group, r_{ip} is the pooled correlation for the i th item, and a is the number of items in a set ($a = 7$ for quality of life, 7 for depression, 36 for stress). If $|d.| \geq .10$, the group was accepted as demonstrating a large enough effect size to be considered an important moderator.³

G. Attention shifted to the possibility that the cumulative statistics might not accurately represent the data. This situation would arise if outlier items were present that inflated the difference scores and were the basis for large χ^2 values. To guard against this possibility, the results for individual items were examined to identify cases where the cumulative statistics might be misleading. The first step in this analysis was to determine whether any items had exceptionally large deviations from the pooled correlations. Items with $|z| \geq 2.50$ were flagged as possible outliers. This screening criterion identifies items with a two-tailed $p < .001$ significance level. This item significance criterion applied Bonferroni adjustments to maintain an experiment-wide error of $p < .05$ with an implied 50 significance tests.⁴ The second step consisted of examining the z -scores of the remaining items to determine whether there was a large gap between the potential outliers and the other items. The average was accepted as representative of the full set of items if the potential outliers represented extremes of a distribution that was generally shifted toward positive or negative scores.

Important Moderators. Statistical tests for significance do not, in themselves, guarantee results that are of theoretical or practical importance (Cohen, 1988). Note that for any group within a moderator analysis the size of the z -scores for individual items depends on both the size of the difference between the correlation observed in that group and the average correlation and the size of the sample in that group. Large z -scores can mean that the difference is large, the sample size is large, or both. The overall χ^2 is the sum of the squared z -scores, so this statistical test adds the possibility that the cumula-

tive significance test is the result of one or two outlier items. The most interesting moderator effect is that which involves a reasonably large effect size that cannot be discounted as chance and that is not confined to just one or two aspects of shipboard life. The presentation of results, therefore, focuses on important moderator effects defined as those for which $|d| \geq .10$, the cumulative χ^2 was significant ($p < .05$), and outliers were not the primary basis for the cumulative χ^2 . The term *important moderator effect* is used to distinguish this multifaceted interpretive criterion from a simple statistical significance test.

RESULTS

Gender Effects

The first analysis step converted gender differences in stress and strain to r_{pb} values. This step expressed the differences in a metric suitable for comparison to the results of previous reviews. As predicted, the full sample produced a consistent, but weak, tendency for women to report higher stress and strain than men (Table 1). Keeping in mind that the scoring for QoL and coping items was reversed so that a high score corresponded to high stress and strain, women reported more stress/strain on 45 of 50 items.

The data also confirmed the prediction that the typical gender difference would be small. The median correlation was $r = .04$. This median correlation translates to 52% of females reporting above average stress versus 48% of males (i.e., $.52 - .48 = .04$; cf., Rosenthal & Rubin, 1982). This median value was very close to the crude estimate of $r_{pb} = .03$ derived in Appendix A.

The analyses failed to confirm the prediction that depression items would be an exception to the general trend. The unweighted average of the correlations for depression items ($r_{pb} = .07$) was larger than those for items pertaining to quality of life ($r_{pb} = -.03$) and stressful conditions aboard ship ($r_{pb} = .04$), but still less than the minimum value for a theoretically or practically important difference (Cohen, 1988).

The analyses supported the prediction that work-family issues would be a major source of gender differences. The two items that produced the largest differences dealt with concerns about children's well-being during deployment ($r_{pb} = .20$) and child care arrangements ($r_{pb} = .25$). These differences were the only ones that substantially exceeded the minimum value ($|r_{pb}| = .10$) for an important effect. Overall, only 8 of 50 items met this minimum criterion for a small effect size. Six of the eight items exceeded the criterion by .03 or less ($|r_{pb}| \leq .13$).

When the correlations were translated into BESDs, women were about half again as likely to be concerned about children and child care as men: Child care, $r_{pb} = .25$, BESD = 62.5% vs. 37.5%; children's well-being during deployment, $r_{pb} = .20$, BESD = 60% vs. 40%). However, women were not extraordinarily concerned about all family issues. Women were only slightly more likely to be concerned about children's discipline than men ($r_{pb} = .08$, BESD = 54% vs. 46%). This effect was below the cutoff for an important effect size, but still was larger than the typical difference in this study. Thus, women generally demonstrated greater concern for children with particular emphasis on child care and well-being.

Table 1. Stress/Strain Correlations to Gender and Navy Variables.

	<u>Gender</u>
<i>Quality of Life</i>	
Job	-.01
Personal life	.00
Health	-.11
Life as a whole	-.02
Family	.00
Spouse	-.11
Children	.07
<i>Depression</i>	
Not get going	.07
Sad	.13
Sleep	.07
Effort	.04
Lonely	.05
Blues	.08
Concentration	.04
<i>Stress</i>	
Life as whole	.05
Financial	-.03
Personal health	.12
Family health	.01
Being aboard	.04
Crowding	.03
Safety	.03
Hygiene	.02
Privacy	.09
Exercise	.06
Recreation	.02
Nutrition	.06
Supervisor	.02
Peers	.07
Subordinates	-.02
Procedures	.04
Living space	.09
Duties	.10
Promotion	.00
Downsizing	-.01
Special friend	.01
Breaking up	.00
Communications	.05
Postdeployment	.02
Confined	.05
Future	.06
Alcohol	-.12
Isolation	.03
Out of touch	.02
Life	.07
Children	.20
Discipline	.08
Child-care	.25
<i>Effect on:</i>	
Personal life	.04
Performance	.01
Coping	-.02

Note. Entries are point-biserial correlations between item responses and gender. Those $|r_{pb}| \geq .10$ are shown in bold typeface.

Table 2. Summary of Tests for Moderator Effects

Moderator	Item Set:			
	Quality of Life	Depression	Stress	Total
Ship Type	$\chi^2 = 86.74$ df = 56 p = .006	79.34 56 .022	363.29 288 .002	489.37 400 .002
Rating	$\chi^2 = 160.83$ df = 161 p = .508	175.74 161 .202	959.59 828 .001	1296.16 1150 .002
Deployment Status	$\chi^2 = 8.32$ df = 7 p = .305	4.47 7 .726	32.34 36 .644	45.12 50 .669
Deployment Experience	$\chi^2 = 30.38$ df = 35 p = .692	25.84 35 .871	165.56 180 .773	221.78 250 .901

Note. "df" is the number of degrees of freedom for the χ^2 statistic. Statistically significant ($p < .05$) χ^2 values shown in boldface.

Moderator Effects

Overview. Tests for moderator effects indicated that ship type and rating influenced the size of gender differences in stress and strain (Table 2). Deployment status indicators did not affect these differences enough to rule out chance as an explanation for the observed variations in the correlations.

Ship Type.⁵ The follow-up assessment of ship type moderator effects indicated that those effects were limited in scope. Each ship type that produced a significant moderator effect did so because it influenced gender differences in exactly one of the three stress/strain indicator categories.

Stress. Only two ship types produced important moderator effects (Table 3).⁶ The largest effect was the trend for women aboard Salvage Ships (ARS). This effect reversed the general tendency for women to report more stress (ARS average $r = -.11$, Pooled average $r = .05$, average $d = -.16$). The trend generalized across the set of items, so the average effect size provided a reasonable summary of the data.⁷ The other important moderator effect was obtained for women aboard Miscellaneous Command Ships (AGF). These women amplified the general trend toward higher stress among females (average $r = .17$, average $d = .12$).

Depression. There were no important depression moderator effects.⁸

Table 3. Effect of Ship Type on Gender Differences in Stress and Strain

	Quality of Life		Depression		Stress	
	<i>d</i>	χ^2	<i>d</i>	χ^2	<i>d</i>	χ^2
ARS	.01	5.93	-.10	8.65	-.16	100.17*
AGE	-.05	1.06	.10	6.73	.12	52.89*
LSD	.28	15.87*	-.15	9.29	-.01	23.08
AE	-.10	15.24*	-.01	2.31	.01	25.47
AO	-.10	15.86*	.07	10.95	.00	14.86
AD	-.01	5.08	.06	17.63*	.04	53.48*
AS	.04	16.57*	-.04	15.45*	-.02	31.79
AS(1)	.05	19.85*	-.06	23.28*	-.03	41.44

Note. Numbers in boldface met the minimum effect size criterion of $|d| \geq .10$ and produced a significant ($p < .05$) χ^2 . These moderator effects, therefore, were both important and statistically significant. The row labeled "AS(1)" is the effect for the Submarine Tenders dropping one outlier submarine (cf., Appendix C).

Quality of Life. Important moderator effects for quality of life were evident for two ship types.⁹ Although men and women reported similar quality of life evaluations in the full sample (average $r = -.02$; Table 1), women aboard Amphibious Dock Landing Ships (LSD) reported substantially better quality of life than their male counterparts (average $r = .26$; average $d = .28$). In contrast, women assigned to Ammunition Ships (AE) reported poorer quality of life than their male counterparts (average $r = -.13$; average $d = -.10$).

Rating Moderator Effects. Nine of 24 ratings produced moderator effects for at least one category of stress and strain indicators (Table 3). The moderator effects were somewhat more general than those for ship type, as three of the nine ratings produced moderator effects for two of the three categories of stress and strain indicators. However, no rating produced moderator effects for all three.

Stress. Important stress moderator effects were noted for three ratings.¹⁰ The largest effect was that women in the Signalman (SM) rating reversed the typical gender effects for stress (average $r = -.18$; average $d = -.23$). The second largest effect was a similar reversal for women in the Quartermaster (QM) rating (average $r = -.14$; average $d = -.19$). The third largest effect was a tendency for women in the Mess Management Specialist (MS) rating to magnify the common tendency for women to report higher stress (average $r = .17$; average $d = .12$).¹¹

Depression. Three ratings produced moderator effects for depression.¹² The two strongest effects reversed the general tendency for women to report higher depression. The average deviation from the pooled trend (average $d = -.23$) was substantial for the Signalman (SM) rating. A similar general tendency was noted for the Torpedo-

man's Mate (TM) rating (average $d = -.18$). Only one rating showed

Table 4. Rating as a Moderator of Gender Differences.

	Quality of Life		Depression		Stress	
	d	χ^2	d	χ^2	d	χ^2
SM	-.05	2.32	-.23	14.77*	-.23	78.35*
QM	.12	8.60	-.16	6.57	-.19	53.50*
MS	-.11	8.99	.10	11.90*	.12	75.24*
DC	.06	9.35	-.12	8.22	-.10	61.86*
TM	.18	17.71*	-.18	15.14*	-.09	34.38
SK	-.04	8.52	.15	16.99*	.07	50.84
SN	.00	4.90	-.06	15.42*	-.04	37.39
MM	.12	19.32*	-.01	7.08	-.02	27.21
HT	-.04	8.99	.03	7.13	.00	59.01*

Note. Numbers in boldface met the minimum effect size criterion of $|d| > .10$ and produced a significant ($p < .05$) χ^2 . These moderator effects, therefore, were both important and statistically significant.

a general moderator effect amplified the typical gender difference in depression. The Storekeeper (SK) rating produced a reliable trend toward amplified depression differences (average $d = .15$).¹³

Quality of Life. Only one rating produced an important quality of life effect. The Torpedoman's Mate (TM) rating demonstrated an overall tendency to reverse the usual observation that women gave lower QoL ratings than men. However, this finding was limited primarily to two items, job QoL ($z = 2.76$) and health and physical condition QoL ($z = 2.49$). The cumulative trend for the remaining items was statistically nonsignificant ($\chi^2 = 3.79$, 5 df, $p < .581$).

Moderator effects were obtained only for stress and depression. The effects involved 5 of 24 ratings, but only a single rating showed any tendency toward a generalized moderator effect encompassing both stressful conditions and strain. Even that instance did not represent an exacerbation of stress and strain. Among Signalmen (SM), women actually reversed the general trend toward higher stress and strain found in the pooled data.

DISCUSSION

Restating the gender contrasts in the WANS study as effect sizes highlights the fact that those differences are, on the average, quite small. In general, gender differences in stress in the U.S. Navy shipboard population are too small to be of theoretical or practical significance. This observation does not make the shipboard population an isolated example of gender equivalence. The differences are, in fact, very similar to those found in civilian populations. Further, the evidence provides little reason to believe that there are isolated

subgroups of women within the shipboard population that deviate markedly from this general conclusion. The moderator analyses identified a few scattered instances of important deviations from the population trends, but these differences were somewhat more likely to indicate lower stress among women than they were to indicate higher stress.

General Trends

The similarity between the gender differences in the shipboard and civilian populations was evident when all stress and strain indicators were considered together. The median gender difference for the shipboard population translated to an effect size ($r_{pb} = .05$). This value was virtually identical to that derived by reviewing an extensive sample of recent studies of civilians ($r_{pb} = .03$; cf., Appendix A).

The similarity between this shipboard population and civilian populations also was evident at finer levels of analysis. The ordering of effect sizes for gender differences in civilian studies was work-family conflict > affective stress > event/condition stress (cf., Appendix A). If concerns about family are equated with work-family conflict, the same ordering was observed in the shipboard population.

Closer comparison of the shipboard and civilian population gender differences in stress suggests that those differences actually may be slightly smaller in the shipboard population. This suggestion is based on two observations. First, shipboard gender differences in depression indicators were smaller than those typically found in studies of civilians ($r_{pb} = .07$ and $r_{pb} \approx .11$, respectively). The difference was slight, but was enough to shift the association from the important effect category (i.e., $r_{pb} > .10$) to the unimportant effect category. Second, only 16% of the shipboard gender differences were large enough to be considered important ($|r_{pb}| \geq .10$) compared to 35% of the gender differences reported for samples of civilians (cf., Appendix A).

The design of this study may account for the trend toward smaller shipboard gender differences in stress. This study compared women to men matched on a number of attributes that might affect stress and strain (e.g., marital status, employment). This matching may have eliminated some sources of differences in depression and stress. Also, the analysis was conducted with individual items. Items generally will be less reliable than multiitem scales. The civilian research included a number of studies that used scales as stress/strain indicators. The higher reliability of the scales will produce stronger correlations between stress and strain and other variables, including gender (Nunnally & Bernstein, 1994). The use of scales, therefore, would be expected to result in a slight increase in the proportion of noteworthy effects. Finally, civilian studies with large samples tended to produce smaller estimated gender differences (Appendix A). The sample size in this study was large relative to the average sample in the civilian studies, so it could have been expected to produce somewhat smaller than average gender differences.¹⁴

Moderator Effects

There was little evidence that there are subgroups of shipboard

women who are under particularly heavy stress. Elevations of both stress and strain are required to infer meaningful elevations of stress. This statement is reasonable if one adopts the view that conditions that do not evoke a substantial response arguably are not stressful. Applying this rule, occupation (rating) was the only organizational factor that moderated stress. Only occupation produced statistically significant moderator effects for both stress and at least one category of strain indicator. Even for occupation, the strain effect was limited to depression.

The picture is slightly different if the proposed "stress + strain" definition of a meaningful moderator is applied to individual groups. The common tendency for women to report higher stress and strain than men was amplified in only the Mess and Maintenance Specialist (MS) rating (stress, ES = .12; depression, ES = .10; QoL, ES = -.11). For ship types, gender differences on stress indicators were exacerbated aboard Miscellaneous Command Ships (AGF; ES = .12). This statistically significant trend was accompanied by a tendency toward higher depression that met the effect size criterion, but not the statistical significance criterion. Thus, only 1, or at most 2, of the 33 subgroups demonstrated amplified gender differences in stress.¹⁵

The fact that one or two groups met the proposed criteria for a meaningful moderator effects must be weighed against the evidence that some groups reversed the general tendency for women to report higher stress and strain. Women in the Quartermaster (QM) rating reported lower stress and strain than their male counterparts (stress, ES = -.19; depression, ES = -.16, QoL, ES = .12). Women in the Signalman (SM) rating also reported lower stress (ES = -.23) and lower depression (ES = -.23) than their male counterparts. For ship types, women assigned to repair ships (ARS) reported lower stress (ES = -.16) accompanied by a depression trend that met the effect size criterion, but not the statistical significance criterion. Thus, if one or two groups amplified the typical gender differences in stress and strain, two or three groups reversed those differences. Here again, there was no general trend toward higher stress and strain in women.

Conclusions

Gender differences are small on the average and are influenced only slightly by organizational factors. These conclusions must be interpreted in light of possible effects of specific elements of the study design and analysis procedures. The possible influence of using individual items as the dependent variables in the analyses, the large sample size, and of matching women and men on factors that may affect stress and strain have been mentioned above in connection with the depression results. Other potential limitations include the possibility that the wrong moderators were chosen for analysis. Work-family conflict is a potential problem for any parent, regardless of ship assignment or occupation. The difference between men and women may be exacerbated in women who are separated or divorced (Roxburgh, 1996; Simon & Corbett, 1996; Thoits, 1987). The effects of family factors (e.g., marital status, age of children), therefore, may be more important than the moderators explored in this study. It also should be remembered that these results represent a snapshot taken at a specific point in time. The integration of women into ships' crews is an ongoing process. On a

ship, stress levels may be affected by recent deployments, rate of crew turnover, ship's policies, manning levels, and other factors. Changes in any of these factors could modify the stress profile for the ship and in doing so might change the size of gender differences in stress and strain. Finally, it is important to keep in mind that the present findings could be due to chance. Although reasonable statistical criteria were employed to identify important effects in this study, different criteria could have been employed that would have produced different results. For example, if adjustments had been made to allow for the number of significance tests implied in the moderator analyses, it is likely that none of the effects would have been identified as meaningful. However, that approach would have suffered from low statistical power because of the small to moderate sample sizes available for the analyses. The present statistical criteria provided a reasonable trade-off between these alternatives, particularly in light of the fact that this study was an initial exploration of the potential moderator effects. Ultimately, however, the point is that statistical criteria are no substitute for replication. None of the findings should be given undue weight until replicated.

The two major conclusions from this study support what Eagly (1995, 1996) calls an "empirical minimalist" position. Women and men aboard ship report very similar levels of stress, so it is more appropriate to emphasize their similarity than to infer differences. However, substantial differences exist in some areas, notably family concerns. Those isolated differences may be more important in some ways than the general equality of women and men. Even minor differences can have substantial impact if they become a factor in repetitive selection decisions (Martell, Lane, & Emmrich, 1996). Family concerns could be a factor in selection processes if, for example, they influence the decision of whether to reenlist. Both this study and the civilian literature reviewed in Appendix A support the view that further study of gender differences in shipboard stress will be more productive if it focuses on specific issues such as family concerns, rather than focusing on general topics such as those pursued in this initial exploratory study.

FOOTNOTES

¹Two ship types represented in the data, Destroyers (DD) and Amphibious Command Ships (LCC), were not included in the analysis. Each of these ship types was represented by a single ship. This fact made it impossible to determine whether any deviations from the trends in the pooled data were really a general characteristic of the type or just a peculiarity of the specific ship studied.

²The same relationships often are expressed as mean differences. A t-test between means then is used to assess statistical significance. These correlations are related to this typical procedure as follows:

$$t = \frac{r_{pb} \sqrt{(n-2)}}{\sqrt{(1-r_{pb}^2)}}$$

where t is the value of the t -test, r_{pb} is the point biserial correlation, and n is the sample size. Thus, t increases as the sample size, n , and as the effect size, r_{pb} , increases. Because some n s were large, small effect sizes potentially would yield statistically significant effects. There was a risk, therefore, that results would be reported as significant when they were of little or no practical or theoretical significance. Reporting r_{pb} focused attention directly on the effect size.

³The correlation corresponding to z_i used to compute V is not necessarily equal to the r_{if} used to compute the $|d|$ statistics. The weighted average of the within-group correlations can differ from the full sample correlation. A difference will occur whenever group means correlate differently than individual scores within groups. Differences also can occur when part of the sample must be omitted from the analysis as in the analysis of ship types. The $|d|$ statistics were computed using the full sample correlations to provide a common reference point for all of the analyses. That reference point is the best estimate of the gender-stress correlations if the null hypothesis of no moderator effects is true. Examination of the pooled within-group correlations indicated these were quite close to the full sample correlations in most cases, so in general $r_{i.} \approx r_{if}$.

⁴Using this criterion, all items were subjected to the same significance test. The test was conservative and kept the implied probability level for the full set of stress and strain items at $p < .05$ for the group.

⁵Preliminary analyses established that ships were reasonably homogenous within ship type (see Appendix B).

⁶Women aboard Destroyer Tenders (AD) produced results that combined statistical significance with a small effect size. The average correlation for these women represented an amplification of the tendency for women to report more stress than men (average $r = .08$; $|d| = .04$). The trend was general across the stress items, but the effect size was small whether expressed as the average size of the correlations or as the difference between the group correlations and the pooled correlations. The statistical significance of the finding was the result of a large sample size ($n = 753$).

⁷Two items exceeded the criterion for identifying an item that individually deviated significantly from the pooled gender difference (life as a whole, $z = -2.90$; people I work with (my peers), $z = -2.51$). The use of an average to represent the moderator effect still was reasonable because these items were not obvious outliers. Three items produced results that approached but did not exceed the $|z| > 2.50$ criterion (relationship with spouse or boyfriend, $z = -2.48$; personal future, $z = -2.48$; feeling isolated and excluded, $z = -2.49$). Another seven items produced z-scores between $z = -2.00$ and $z = -2.48$. Fifteen more items produced z-scores between $z = -1.00$ and $z = -2.00$. Thus, the two significant items were part of a distribution in which 27 of 36 items had z-scores of -1.00 or less.

⁸Women aboard Destroyer Tenders (AD) amplified the general tendency for women to report more depressive symptoms (average $r = .13$; $|d| = .06$), while women aboard Submarine Tenders (AS) showed essentially no difference from men (average $r = .03$; $|d| = -.04$). In both cases, the small differences were significant because of a large sample. None of the individual items met the item significance criterion ($|z| < 2.14$) for Destroyer Tenders, so the average effects were representative of the full set of items for depression and stress. The largest z-score for an item in the Submarine Tender sample was $z = -2.74$. Given the sample size ($n = 1,144$), this z-score corresponds to a difference of .081 between the correlation observed in Submarine Tenders and the pooled correlation for all ship types. Thus, even the maximum difference for this ship type would not meet the minimum effect size standard set for identifying moderators in this study.

⁹Two other ship types produced statistically significant moderator effects for quality of life, but failed to meet other criteria for an important moderator effect. First, Submarine Tenders (AS) produced statistically significant effects for quality of life ($|d| = .04$, $\chi^2 = 16.57$), but did not meet the effect size criterion. Excluding the one clear outlier for this ship type (cf., Appendix C) increased the effect sizes slightly (quality of life, $|d| = .05$, $\chi^2 = 19.85$), but the effect remained well below the minimum effect size criterion. These results, therefore, were due primarily to the relatively large sample size for this ship type than to a substantial deviation of the ship type from the profile for the average ship. Oilers (A) also produced statistically significant moderator effect, but in this case a single item was the primary basis for statistical significance. The item for life as a whole ($z = -2.70$) accounted for 46% of the overall χ^2 ($\chi^2 = 7.29$). The remaining six items did not deviate significantly from the pooled correlations ($\chi^2 = 8.57$, 6 df, $p < .200$). Whether the "life as a whole" item was a true outlier is uncertain. All of the items produced negative z-scores ($-2.11 < z < -.12$). However, the largest z-score for the remaining items was substantially less than that for the life as a whole item, and the importance of a moderator effect that disappears with the removal of a single item is questionable.

¹⁰Two ratings produced significant, but narrowly defined stress effects. Hull Technicians (HT) reversed the typical tendency for women to report greater gender lack of privacy aboard ship ($z = -2.52$), greater inability to get enough exercise aboard ship ($z = -2.65$), and greater stress from people with whom living space was shared ($z = -3.00$). A similar reversal for lack of recreational activities approached significance ($z = -2.35$). No other item had a $|z| > 2.00$. The average effect for the

stress items was zero because 16 of the 32 remaining stress items had positive z-scores. Thus, the statistically significant overall trend was focused on a few specific stressors. Females in the Damage Controlman (DC) rating reported less stress than men, primarily because they felt less stress related to discipline of children ($z = -3.20$) and child care arrangements ($z = -3.07$). Removing these two items, the remaining stress items showed no evidence of a moderator effect ($\chi^2 = 42.20$, 34 df, $p < .158$).

¹¹The Quartermaster (QM) rating produced two z-scores for stress that met the item significance criterion (child care arrangements, $z = -3.01$; relationship to spouse or boyfriend/girlfriend, $z = -2.51$). While those z-scores were large, they did not stand out from the overall distribution of z-scores for this rating. Two other items produced z-scores in the -2.00 to -2.50 range, and 10 items produced z-scores in the -1.00 to -2.00 range. The largest positive z-score was $z = .63$. This pattern suggests that the average difference was representative of the rating. In the Mess Management Specialist (MS) rating, the stress items dealing with ability to communicate with friends and family ($z = 2.73$) and discipline of children ($z = 2.59$) met the item significance criterion. However, six other items produced z-values between 2.00 and 2.50 and 11 items produced z-values between 1.00 and 2.00. Thirty-four of 36 z-scores were positive (34 of 36), and the largest negative z-score was only $z = -.27$. For Signalmen (SM), one item (stress related to personal future, $z = -3.40$) exceeded the item significance criterion level, but 35 of 36 z-scores were negative. The negative z-scores tended to be moderate in size as 4 were between -2.00 and -3.00 and 18 were between -1.00 and -2.00. The personal future item may be an outlier in these data, but it is part of a trend that is less pronounced for other items. To the extent that it is an outlier, including the personal future item may slightly distort the overall estimated value of the moderator effect for this occupation, but the distortion should be slight given that this is only 1 of 36 items.

¹²The depression trend for the Mess Management Specialist (MS) rating was statistically significant, but barely met the effect size criterion (average $d = .10$). Further, the average difference was attributable largely to the difference for the single item "couldn't get going" ($z = 2.92$). Removing this item, the χ^2 for the remaining six items was clearly nonsignificant ($\chi^2 = 3.37$, 6 df, $p < .762$). Among Damage Controlmen (DC), the depression effect was large enough to meet the effect size criterion and was stable across items. However, in this case, the sample size was too small to make the cumulative χ^2 or any single item deviation statistically significant ($|z| < 1.80$ for all items).

¹³For Storekeepers (SK), the z-scores for individual depression items ranged from 0.97 to 1.93, thereby indicating a consistent effect across the set of items. For Signalmen (SM), z-scores for these items ranged from $z = -0.30$ to $z = -2.08$. For Torpedoman's Mate (TM), z-scores ranged from $z = -2.16$ to $z = 0.45$.

¹⁴The association between sample size and effect size may be the product of publication biases. Larger samples mean that smaller effects will pass significance tests (Rosenthal & Rosnow, 1984). If studies are published only when statistically significant effects are present, studies with small sample sizes will be published only when they produce

large effects. The observed association between sample size and effect size would arise naturally if all studies that produced significant differences between men and women were published. This line of argument buttresses the conclusion that true gender differences are small because large sample sizes will provide more precise estimates of the true effect sizes than will small sample sizes.

¹⁵Although there were no specific hypotheses regarding their effects, the study also examined ship size and crew composition as factors that might influence the size of gender differences in stress. Toward this end, moderator analyses were performed with ship crew size and the proportion of females aboard ship as the grouping variables.

The study design sampled all women aboard ship, plus a subset of men matched to those women on specific characteristics, such as age and occupation (Garland & Timberlake, 1995). The proportions of men and women in the sample, therefore, were not representative of the proportions in the overall shipboard populations. Information on the number of men and women in each crew were taken from a combination of computerized records maintained at the Naval Health Research Center and information provided by the personnel department of each ship studied (Garland & Timberlake, 1995). This information was used to determine the overall size of the ship's population and the proportions of men and women aboard ship.

The number of people assigned to the ship did not affect the size of gender differences in stress or strain (Quality of Life, $\chi^2 = 38.43$, 28 df, $p < .091$; Depression, $\chi^2 = 15.15$, 28 df, $p < .977$; Stress, $\chi^2 = 148.62$, 144 df, $p < .379$). The proportion of women aboard ship did not affect quality of life ($\chi^2 = 34.68$, 28 df, $p < .189$) or depression ($\chi^2 = 16.69$, 28 df, $p < .955$), but did affect stress levels ($\chi^2 = 223.53$, 144 df, $p < .001$). Follow-up examination of the basis for the statistically significant effect of proportion of women aboard ship showed the significant deviations from the full sample gender differences were as follows:

Ships with 19%-26% Women. Women aboard these ships amplified the general trend toward slightly higher stress in women. The difference was larger than average for crowding ($r_{pb} = .16$; $d = .13$), hygiene ($r_{pb} = .13$; $d = .11$), privacy ($r_{pb} = .20$; $d = .11$), living space ($r_{pb} = .24$; $d = .16$), and promotion opportunities ($r_{pb} = .13$; $d = .12$). The trend toward higher stress aboard these ships did not apply to all areas of stress. Women aboard these ships showed no tendency to feel more stressed by concerns about children ($r_{pb} = .00$; $d = -.21$) and child-care arrangements ($r_{pb} = .02$; $d = -.24$).

Ships with 27%-29% Women. In several areas, this group reversed the general tendency for women to report more stress. However, the deviations from the general trends were all small (family health, $r_{pb} = -.04$, $d = .08$; being aboard ship, $r_{pb} = -.06$, $d = .09$; peers, $r_{pb} = .00$, $d = -.07$; typical procedures, $r_{pb} = -.04$, $d = -.08$; living space, $r_{pb} = -.02$, $d = -.10$; performance, $r_{pb} = .01$, $d = -.09$; promotion, $r_{pb} = -.10$, $d = -.11$; feeling confined, $r_{pb} = -.03$, $d = -.08$; isolation, $r_{pb} = -.07$, $d = -.10$). Only the two largest differences met Cohen's minimum criterion of $|d| \geq .10$, so the average effect size was below that minimum.

Taken together, the results showed that only ships with 17%-26% females in the crew produced an important moderator effect. Aboard those ships, stresses related to living conditions were more pronounced for females than males. This effect may occur because the requirement for separate berthing areas leads to more crowding for women. For example, if a ship has two berthing areas for women, each area might be designed to house 25 sailors. If 50 women are assigned to the ship, these two areas may be completely full even though there are vacancies in berthing areas for males. This uneven distribution may be necessary because the only other option is to move all of the males out of a berthing area designed for 50 or even 100 sailors. This movement may not be feasible within the constraints of having a ship that is near its full crew complement. Note, however, that whatever the reason for the habitability stress differences, the trend did not generalize to other areas of life aboard ship. The reversals noted for concerns about children are evidence that there was no general effect of proportion of females on gender differences in stress.

¹⁶The remaining studies stated only that results were "nonsignificant."

¹⁷This figure is inflated by a single sample of 13,017, but the median sample size was 286; 25% of the samples had sample sizes of $n \geq 778$.

¹⁸Five differences were listed only as "significant." Their location in the distribution is uncertain. However, the omission of less than 1% of the total number of correlations could not possibly shift the distribution dramatically.

¹⁹Five studies combined two or more approaches (e.g., by providing means and standard deviations for all stress indicators with t-tests for those that demonstrated statistically significant differences). Those studies have been classified here in terms of the method used to extract the majority of the effect size estimates.

²⁰Considering the combined effect of two trends illustrates the potential problem. The two trends involved the effects of having multiple indicator variables and the effects of sample size.

A. Half of the studies reviewed here employed five or more stress indicators. Given five indicators, the probability that at least one test will be statistically significant by chance is $p = .226$. This probability is substantially higher than the $p < .05$ implied for a single significance test. Applying the directional hypotheses used in this study, the net result would be that 11.3% of the studies would produce at least one nominally significant effect. The actual rate will be higher because many studies involved more than 5 indicators.

B. Sample size also can contribute to the illusion that substantial stress effects are present. One fourth of the studies reviewed had 778 or more participants. For those studies, a correlation of $r_{pb} = .07$ was large enough to be statistically significant ($p < .05$). However, many "significant" findings will be unimportant. If $r_{pb} = .10$ is the lower bound for an important finding, this lower bound corresponds to $p < .003$. In this exam-

ple, 94% (i.e., .047/.050) of the nominal "statistically significant" results would be unimportant from a theoretical and practical perspective. Thus, for nearly 25% of the studies, there is a substantial probability that a result can be statistically significant, but unimportant.

Sample size and number of variables studied were not very strongly related in the studies reviewed ($r = .03$; $r = .16$ with largest sample excluded). The implication is that the effects of having multiple stress indicators and large sample sizes would be additive. Applying elementary probability theory, suppose the effects of having a large number of variables and the effects of having a large sample are independent. If so,

$$\begin{aligned} p(a \text{ or } b) &= p(a) + p(b) - [p(a) * p(b)] \\ &= .113 + .250 - (.113 * .250) \\ &= .335 \end{aligned}$$

Thus, roughly one study in three would be expected to produce a nominally important result that either was attributable to chance or to a large sample size.

²¹The general formula for the relationship between true population correlations, indicator reliability, and observed correlations is:

$$r_{xy} = \rho_{xy} * \sqrt{(r_{xx} * r_{yy})}$$

where r_{xy} is the observed sample correlation, ρ_{xy} is the true population correlation, r_{xx} is the reliability of one variable, and r_{yy} is the reliability of the other variable. Substituting $r_{xy} = .05$, $\rho_{xy} = .30$, $r_{yy} = 1.00$ (for gender, a construct that presumably can be measured unambiguously), and solving for r_{xx} yields $r_{xx} = .41$.

²²Gender differences were compared with the two large positive categories from Table A-2 combined. The distribution for the studies with large sample sizes differed substantially from that for the studies with small sample sizes ($\chi^2 = 54.61$, 4 df, $p < .0001$). The differences may have been influenced by the fact that a larger proportion of the associations were simply reported as nonsignificant in the smaller samples (35.3% vs 13.3%). However, the presence of more extreme effect sizes would be expected as a natural consequence of greater sampling variation in smaller samples. The important point is that the present analyses are more likely to be biased toward overestimating gender differences rather than underestimating them.

²³Gender differences for studies with six or fewer stress indicators were compared to those for studies with more than six indicators. The two large positive categories from Table A-2 were combined for the comparison. The two distributions differed substantially ($\chi^2 = 50.84$, 4 df, $p < .0001$). The frequency of trivial differences was greater in studies with large numbers of indicators (74.8% vs 42.2%). The former percentage reflects 223 gender differences; the latter reflects 31 gender differences. Averaging the results from those studies, therefore, would have reduced the proportion of total correlations falling in the $|r| \leq .10$ range. Note, however, that nearly 50% of the differences fell in the $|r|$

$\leq .10$ range even when attention was restricted to studies with only a few indicators.

REFERENCES

Bedeian, A. G., Burke, B. G., & Moffett, R. G. (1988). Outcomes of work-family conflict among married male and female professionals. *Journal of Management*, 14, 475-491.

Beutell, N. J., & O'Hare, M. M. (1987). Work-nonwork conflict among MBAs: sex differences in role stressors and life satisfaction. *Work & Stress*, 1, 35-41.

Brown, D. R., & Gary, L. E. (1987). Stressful life events, social support networks, and the physical and mental health of urban Black adults. *Journal of Human Stress*, 13, 165-174.

Caldwell, R. A., Pearson, J. L., & Chin, R. J., (1987). Stress-moderating effects: social support in the context of gender and locus of control. *Personality and Social Psychology Bulletin*, 13, 5-17.

Caplan, R. D., Cobb, S., French, J. R. P., Jr., Harrison, R. V., & Pinneau, S. R., Jr. (1975). Job demands and worker health (HEW Publication No. (NIOSH) 75-160). Washington, DC: U.S. Government Printing Office.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). NY: Academic Press, 1988.

Conger, R. D., Lorenz, F. O., Elder, G. H., Simons, R. L., & Ge, X. (1993). Husband and wife differences in response to undesirable life events. *Journal of Health and Social Behavior*, 34, 71-88.

Cook, M., Young, A., Taylor, D. & Bedford, A. P. (1996). Personality correlates of psychological distress. *Personality and Individual Differences*, 20, 313-319.

Cushway, D. (1992). Stress in clinical psychology trainees. *British Journal of Clinical Psychology*, 31, 169-179.

Eagly, A. H. (1995). The science and politics of comparing women and men. *American Psychologist*, 50(3), 145-158.

Eagly, A. H. (1996). Differences between women and men: their magnitude, practical importance, and political meaning. *American Psychologist*, 51, 158-159.

Garland, F. (1995). *Women aboard Navy ships* (Year One Report). San Diego, CA: Naval Health Research Center.

Garland, F. C., & Timberlake, D. S. (1995). *Descriptive tables of demographic characteristics, family structure, and women's health-related issues, and occupational exposures of personnel participating in the U.S. Navy Women Aboard Ship Study*. San Diego, Ca: Naval Health Research Center.

Goffman, E. (1961). *Asylums*. Garden City, NY: Anchor Books.

Greenglass, E. R., & Burke, R. J. (1988). Work and family precursors of burnout in teachers: Sex differences. *Sex Roles*, 18, 215-229.

Hays, W. L. (1963). *Statistics for the social sciences*. NY: Holt, Rinehart & Winston.

Hendrix, W. H., Spencer, B. A., & Gibson, G. S. (1994). Organizational and extraorganizational factors affecting stress, employee well-being, and absenteeism for males and females. *Journal of Business and Psychology*, 9, 103-128.

Heppner, P. P., Cook, S. W., Strozier, A. L., & Heppner, M. J. (1991). An investigation of coping styles and gender differences with farmers in career transition. *Journal of Counseling Psychology*, 38, 167-174.

Hinkle, L. E. (1973). The concept of "stress" in the biological

and social sciences. *Social Science and Medicine*, 1, 31-48.

Holmes, T. H., & Rahe, R. H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research*, 11, 213-218.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis*. Beverly Hills, CA: Sage.

Jick, T. D., & Mitz, L. F. (1985). Sex differences in work stress. *Academy of Management Review*, 10, 408-420.

Lester, D., Posner, I., & Leitner, L. A. (1986). Stress scores of workers by sex and age. *Psychological Reports*, 58, 110.

Mandel, D. R., Baron, R. M., & Fisher, J. D. (1980). Room utilization and dimensions of density: effects of height and view. *Environment and Behavior*, 12, 308-319.

Martell, R. F., Lane, D. M., & Emmrich, C. (1996). Male-Female differences: a computer simulation. *American Psychologist*, 51, 157-158.

Martin, J. A., Acevedo, G., & Garland, F. C. (In review). Comparison of men and women aboard Navy ships: Life stress conditions, psychological stress, distress, coping and quality of life issues.

Martocchio, J. J., & O'Leary, A. M. (1989). Sex differences in occupational stress: A meta-analytic review. *Journal of Applied Psychology*, 74, 495-501.

McCreary, D. R., Wong, F. Y., Wiener, W., Carpenter, K. M., et al. (1996). The relationship between masculine gender role stress and psychological adjustment: a question of construct validity? *Sex Roles*, 34, 507-516.

Meehl, P. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 46, 806-834.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). NY: McGraw-Hill.

Pugh, W. M., Gunderson, E. K. E., & Dean, L. M. (1975). Sources of situational variance in environmental perceptions (Tech. Rep. 75-11). San Diego, CA: Naval Health Research Center.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1, 385-401.

Richman, J. A., & Flaherty, J. A. (1990). Gender differences in medical student distress: Contributions of prior socialization and current role-related stress. *Social Sciences and Medicine*, 30, 777-787.

Rosenthal, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.

Rosenthal, R. & Rosnow, R. L. (1984). *Essentials of behavioral research*. NY: McGraw-Hill.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, 74, 166-169.

Roxburgh, S. (1996). Gender differences in work and well-being: Effects of exposure and vulnerability. *Journal of Health and Social Behavior*, 37, 265-277.

Rubin, R. T., Gunderson, E. K. E., & Arthur, R. J. (1971). Life stress and illness patterns in the U.S. Navy. IV. Environmental and demographic variables in relation to illness onset in a battleship's crew. *Journal of Psychosomatic Research*, 15, 89-94.

Salgado de Snyder, V. N., Cervantes, R. C., & Padilla, A. M. (1990). Gender and ethnic differences in psychosocial stress and gener-

alized distress among Hispanics. *Sex Roles*, 22, 441-453.

Sekaran, U. (1986). Significant differences in quality-of-life factors and their correlates: A function of differences in career orientations or gender? *Sex Roles*, 14, 261-279.

Simon, F., & Corbett, C. (1996). Road traffic offending, stress, age, and accident history among male and female drivers. *Ergonomics*, 39, 757-780.

Sowa, C. J., & Lustman, P. J. (1984). Gender differences in rating stressful events, depression, and depressive cognition. *Journal of Clinical Psychology*, 69, 615-622.

Spielberger, C. D., & Reheiser, E. C. (1994). The Job Stress Survey: Measuring gender differences in occupational stress. *Journal of Social Behavior and Personality*, 9, 199-218.

SPSS, Inc. (1992). *SPSS reference guide*. Chicago: SPSS, Inc.

Swanson, V., Power, K., & Simpson, R. (1996). A comparison of stress and job satisfaction in female and male GPs and consultants. *Stress Medicine*, 12, 17-26.

Thoits, P. A. (1987). Gender and marital status differences in control and distress: Common stress versus unique stress explanations. *Journal of Health and Social Behavior*, 28, 7-22.

Turner, R. J., & Avison, W. R. (1989). Gender and depression: Assessing exposure and vulnerability to life events in a chronically strained population. *Journal of Nervous and Mental Disease*, 177, 443-455.

Vickers, R. R., Jr., & Martin, J. A. (In review). *Exploration of stress differences by gender aboard U.S. Navy ships*. San Diego, CA: Naval Health Research Center.

Walden, T. A., Nelson, P. A., & Smith, D. E. (1981). Crowding, privacy, and coping. *Environment and Behavior*, 13, 205-224.

Waelde, L. C., Silvern, L., & Hodges, W. F. (1994). Stressful life events: Moderators of the relationships of gender and gender roles to self-reported depression and suicidality among college students. *Sex Roles*, 30, 1-22.

Wu, X., & DeMaris, A. (1996). Gender and marital status differences in depression: The effects of chronic strains. *Sex Roles*, 34, 299-319.

Zuckerman, D. M. (1989). Stress, self-esteem, and mental health: How does gender make a difference? *Sex Roles*, 20, 429-444.

APPENDIX A

A Review of Recent Studies of Gender Differences in Stress

A substantial body of evidence on gender differences in stress has accumulated since the mid-1980s. During that period, some social trends may have operated to minimize gender differences in stress, others to enhance them. For example, more women probably work outside the home at present than was the case in 1985. The work may provide opportunities for self-fulfillment, reductions in financial stress, and other benefits. At the same time, having a job can lead to stresses that are not present in the homemaker role. Conflict between home and family, risk of job loss, hostile working environments, and similar problems may be more common now than they were when the work reviewed by Martocchio and O'Leary (1989) and Jick and Mitz (1985) was performed. Even in the absence of significant changes in the proportion of women working, movement into nontraditional occupational categories may have shifted the magnitude of stress differences between men and women.

A broad sample of the research performed subsequent to the mid-1980s was reviewed to determine whether the earlier conclusions should be modified in light of that later work. Anticipating the findings that will be presented below, the basic conclusions from the review were that the general trend toward modest, even trivial, gender differences in stress still holds. However, there may be important exceptions to the trend. Conflict between work and family and indicators of psychological strain are notable exceptions to the rule. Methodological factors combined with the common tendency to employ informal nose counting procedures when summarizing data trends may lead to a mistaken impression that large gender differences in stress are common.

Literature Review Procedure

Selection of Studies. The review of research on gender differences in stress was not comprehensive. A search of the PsycLit® data base combining the search terms "human sex differences" and "stress" identified 1,027 articles with both terms. Project resources were not adequate to fully review this body of literature; so several steps were taken to reduce the material to a manageable size:

A. The PsycLit® entries were reviewed to identify the entries that appeared most relevant to the problem at hand. In general, entries that referred to gender or sex differences and either stress, strain, or distress in the title were retained. Entries that appeared to emphasize exposure to experimental manipulations of "stressors" in the laboratory and/or biological stress indices were dropped. The resulting set of 208 entries focused on studies that appeared to involve objectives and methods similar to those in the WANS study.

B. The 208 entries retained from the first step were examined to identify books and book chapters and research dealing with children or young adolescents (i.e., high school age or less). Fifteen book chapters and 13 articles on adolescents were dropped. The

remaining 180 articles were journal articles involving adults and were likely to contain actual data on gender differences.

C. Ninety-nine (99) of the 180 articles were available in the combined libraries of the Naval Health Research Center and the Navy Personnel Research and Development Center. These 99 articles were reviewed to determine which ones actually contained useable data. Seventy-two articles listed at the end of this appendix contained information initially judged suitable for this review.

D. Three articles provided information that, upon further examination, could not be used effectively in this review. Savery (1990) reported a number of comparisons between men and women, but it was impossible to determine how many total comparisons were made. A study by Lewis and Cooper (1987) was dropped despite greater detail in the reporting of results because the presentation still made it impossible to tell how many comparisons actually were made. Gerdes (1995) reported comparisons among three groups, men, employed women, and housewives. The data were not reduced to single degree of freedom significance tests (e.g., men vs employed women), so differences could not be translated into r_{pb} (see below). One final study was dropped because the results were dramatically at odds with other studies, presenting much larger gender differences than the typical study (McIntosh, Keywell, Reifman, & Ellsworth, 1994). That study appeared to be an outlier in the overall data set, perhaps because it focused on stresses specific to the law school setting in which the study was performed. The decision to exclude this study removed five correlations from consideration and would not significantly affect the conclusions drawn from the summary.

The remaining 68 studies provided 71 samples with data suitable for review. There were more samples than studies because some studies contrasted males to two distinct samples of females.

The 68 studies that provided the data for the review should provide a reasonably representative data set. While these studies represent only a subset of the recent literature on gender differences in stress, the studies reviewed were drawn from journals representing diverse fields, including journals specializing in stress, mental health, clinical psychology, environmental psychology, social psychology, occupational psychology, industrial/organizational psychology, and management. This coverage should sample research based on most or all of the major theoretical perspectives on stress. The overall pattern in the findings should not be subject to major biases arising from the perspective or methods of any single research community.

The evidence reviewed has one obvious major possible source of bias. The original literature review identified studies that included human sex differences as a classification term. Many other stress studies may have examined gender differences as a secondary or incidental consideration. For example, these differences might be evaluated merely to determine whether gender must be controlled when testing study hypotheses about sources of stress or the effects of stress. Studies that were concerned primarily with other issues would not be expected to

provide the results of these comparisons, particularly if there were no important differences.

The failure to report on gender differences in stress would be unimportant if there were no reason to believe this omission biased the available evidence on gender differences. It would be unwise to assume that this is the case. Instead, it seems reasonable to assume that gender becomes an explicit topic of investigation whenever preliminary tests indicate the presence of large gender differences. In this case, the unreported differences from "silent" gender studies would mean that studies with small gender differences would be systematically excluded from the summary. This exclusion would bias the summary in the direction of overestimating gender differences. The review findings leave little doubt that such bias, if present, is unimportant. Even if this bias is operative, the data reviewed did not lead to the conclusion that there are substantial general gender differences in stress.

The Magnitude of Gender Differences. Quantitative literature reviews commonly estimate average effect sizes to summarize the overall body of results. A single number such as the average standardized difference between two groups or the average correlation between two indicators is the main product of these reviews (e.g., Martocchio & O'Leary, 1989).

This review did not generate a point estimate of the size of gender differences in stress. Instead, attention focused on describing the distribution of gender differences in stress and strain. This objective was adopted because the available literature was ill-suited to estimating a single average value. Many studies reported only those gender contrasts that were statistically significant. For example, Chaplain (1995) reported *t*-tests for 2 of 18 stress items, while Brown and Fielding (1993) reported *t*-tests for only 8 of 38 stress items. Any averaging procedure would require inserting some estimate of the missing differences into the review. Those estimates necessarily would be imprecise and could become a source of bias themselves. Options for such substitution include setting all of the missing associations to zero or to the largest statistically nonsignificant correlation. The first option would underestimate the degree of association in most cases assuming the distribution of effects is skewed toward positive or negative values. The second option would overestimate the degree of association in most cases. In either case, the estimation error would bias the final overall estimate of the association when the substituted values were combined with the values that actually were reported in the literature. Given the number of associations involved, the arbitrary assumptions required to make such substitutions could affect the results of the summary. It seemed better to avoid such assumptions if possible.

Limitations of the typical information reported about gender differences was another reason for focusing on the distribution of those differences. Most studies reported means and standard deviations and/or significance test statistics (i.e., a *t*-test or *F*-test) with no direct estimate of effect size. Only a minority of studies (7; 10.6%) expressed their findings as effect sizes, specifically the r_{pb} between stress and gender.¹⁶ For reasons discussed in more detail below, effect size was the focus of the present review, so it was necessary to transform the data

from studies that did not report r_{pb} to this form. The desired r_{pb} values were estimated by applying a standard formula (Hays, 1963, p. 529) to reported t -tests values if these statistics were reported. The desired r_{pb} was estimated from reported F -tests contrasting gender by noting that $F = t^2$ when the F -test has a single degree of freedom (Hays, 1963, p. 354). However, the most common situation required preliminary computations because the reported means and standard deviations were available, but not the t - or F -tests. In these cases, t -statistics were computed using a standard formula (Hays, 1963, pp. 315-316), after which the result was inserted into the formula for computing r_{pb} from t -values.

The process of transforming means and standard deviations into point biserial correlations was imprecise. Some studies reported the means and standard deviations for all stress and strain indicators and the t -tests for the statistically significant differences. When the t -values estimated from the mean and standard deviation were compared to the t -values reported in the paper, the results were close, but not identical. The difference implied that the sample estimate of effect size could only be approximated for many of the gender comparisons in the review. The approximations were reasonably good, but slightly inaccurate. Averaging such data to produce a single number in the form of an estimated average correlation seemed likely to contribute to a mistaken impression of precision in the results. Focusing on the distribution of the gender differences avoided this problem, but still provided a reasonable cogent conclusion about that state of knowledge.

A large volume of evidence was covered in this review whether expressed in terms of the total number of gender differences examined (560) or the total sample size for the combined studies ($N = 53,540$).¹⁷ Given this volume of data, the general distribution of gender effects should be adequately defined, provided no hidden bias affects the process used to select the studies reviewed.

Primary Findings

The distribution of gender differences in stress was consistent with prior findings (Table A-1). Large, pervasive gender differences in stress or strain were not evident. Although most (61.7%) comparisons indicated higher stress or strain for women, a sizable minority (38.3%) showed lower strain for women. The distribution of differences shown in Table A-1 makes it clear that the median r_{pb} must be $.00 < r_{pb} < .10$. If r_{pb} values were evenly distributed over this range, the median correlation would be $r_{pb} = .03$. While this estimate differs from Martocchio and O'Leary's (1985) estimate ($r = -.02$), both fall well within Cohen's (1988) boundaries for unimportant results.

Results that were reported merely as "nonsignificant" had the potential to produce a substantial shift in the distribution of gender differences. Suppose that all 129 results reported simply as significant or nonsignificant had been obtained in studies with small sample sizes, had involved higher stress for women than for men, and had approached significance. *Given this improbable scenario, the distribution in A-1 would shift to the right, but the basic conclusions reached would not change.* The conclusions would not change; the median r_{pb} still would be

Table A-1. Distribution of Estimated Effect Sizes for Gender Differences in Stress

<u>Effect Size</u>	<u>Number of Effects</u>	<u>Percentage of Total Effects</u>
Less than $-.10$	38	8.8%
$-.1$ to $.0$	127	29.5%
$.0$ to $.1$	155	36.0%
$.1$ to $.2$	79	18.3%
$.2$ to $.3$	26	6.0%
Greater than $.3$	6	1.4%
Significant	7	
Nonsignificant	122	

Note. Associations that were given simply as "significant" or "nonsignificant" were not used in computing the percentages.

less than $.10$, even if all 122 statistically nonsignificant correlations were $r_{pb} \geq .10$.

Information about the study sample sizes makes it possible to demonstrate that the extreme situation pictured in the preceding paragraph does not hold. Sample size information can be used to compute the largest value of r_{pb} that would be statistically nonsignificant in each study. This maximum was less than $.10$ for 56 differences. Another 19 associations were in the range from $.10 \leq r_{pb} < .15$. Even the remaining 47 differences would be at most $r_{pb} \leq .20$.¹⁸ Thus, further detail on these associations would add at most 66 associations to the $.10 < .20$ range. The median value still would be $r_{pb} < .10$, and the fact that many of the "nonsignificant" values fell in this domain would support the view that the true median value would be close to $r_{pb} = .00$.

Specific Indicators. The distribution of effect sizes could depend on the specific stress indicators used in the studies. Three types of stress/strain indicators appear to produce distributions of gender differences that are larger than those for the majority of stress/strain indicators.

A. *Depression.* Some of the tendency for the distribution of effect sizes to be skewed toward higher values may be the result of including negative affect indicators in the review. Thirteen of 23 gender differences in depression were $r \geq .10$. The median effect size for these differences was $r_{pb} = .13$.

B. *Distress.* Measures of generalized distress or psychological problems such as the General Health Questionnaire (cf., Cook, Young, Taylor, & Bedford, 1996) added eight more positive differences ranging from $r_{pb} = .11$ to $r_{pb} = .32$; one nonsignificant result was reported. In the context of the overall distribution of results, seven of the nine differences were relatively large ($r_{pb} \geq .18$).

C. *Work-Family Conflict*. Work-family conflict is a second area of stress that shows consistent gender differences. In this area, women consistently score higher than men. Twelve of 15 gender differences were $r_{pb} > .10$. The median effect size for these differences was $r_{pb} = .16$.

Taken together, 71% (32 of 45) of the gender differences involving these three indicator categories were in the $r_{pb} > .10$ range. Only 26% (111 of 431) of the gender differences summarized in Table A-1 showed a similarly strong tendency for women to report higher stress. If the 45 associations involving depression, generalized distress, and work-family conflict were removed from the table, only 20% (79 of 386) of the remaining associations would be in the $r_{pb} > .10$ range. The subset of stress/strain indicators has approximately 3.5 times the expected rate of important differences if one adopts the directional hypothesis that women will report more stress than men.

Illusory Gender Differences?

A stereotype that women and men are differentially susceptible to stress seems to be alive and well. If not, how can the substantial recent literature on the topic be explained? This literature continues to grow despite the fact that reviews conducted nearly 15 years ago pointed out that such differences are slight. Common research practices may contribute to the persistence of an illusion of important differences. These practices include:

A. *Results are seldom expressed as effect sizes*. As noted earlier, only 7 of 72 studies summarized here reported correlations. Effect sizes typically had to be computed from means and standard deviations (31 studies) or from significance tests (23 studies). In fact, more studies simply described the gender differences as significant or nonsignificant (10) than reported correlations.¹⁹

B. *Significance tests are conflated with theoretical and practical importance*. The theoretical and practical importance of a result are determined by effect size. The outcome of a significance test results depend on the combination of effect size and sample size (Rosenthal & Rosnow, 1984). A small effect can be statistically significant if the sample is large. In 45.7% of the studies reviewed the sample size was large enough to produce a significant finding even though $r_{pb} < .10$; in 25% of the samples a finding could be significant even though $r_{pb} < .07$. The accumulation of a large number of statistically significant results in the absence of any large effects is a significant concern in this domain.

C. *Isolated differences are interpreted as representative of the study findings*. Most studies involved more than one stress or strain indicator (median = 5). Typically, gender differences for each indicator were tested for significance individually. Any significant differences then were interpreted as evidence that "stress" differed or, in more cautious cases, that some types of stress differed. However, given that the typical study tested five or more stress and/or strain indicators, the probability that at

least one difference would be significant by chance alone was substantial ($p = .226$ for five indicators). Based on the summary provided here, excessive generalization from isolated effects is particularly likely to be a problem if indicators of depression, general distress, or work-family conflict are included in the study.

D. Replication is rarely a concern. Few studies attempted explicitly to replicate findings from prior research. In studies involving two or more samples, results for each sample typically were interpreted as independent findings reflecting the unique characteristics of the different groups (e.g., Sekeran, 1986). This practice has not been followed in all cases, as some studies have tested for interactions between gender and occupation (Spielberger & Reheiser, 1994; Swanson, Power, & Simpson, 1996). The inference that different subgroups of men and women represent groups that should have the status of distinct entities in theoretical formulations may be correct where it has been drawn, but systematic tests of the hypothesis that the groups truly differ would be desirable.

E. Stress often is treated as a broad construct. The very breadth of the concept is useful when it points to commonality of interest across what appear to be diverse research fields (Hinkle, 1973). However, breadth is counterproductive when a focus on more specific topics is appropriate. This case arguably applies to gender differences in stress where depression, general distress, and work-family conflict appear to define distinctive subsets of results within the larger stress domain.

The combined effect of these research practices is a high probability that misleading conclusions will be reached in any single study. The inference that differences occur with greater than chance frequency may be drawn because a substantial proportion of studies show at least one statistically significant gender difference among a set of variables. The significant difference(s) may be too small to be of practical or theoretical importance. Treating stress as a general concept may lead to the inference that two studies qualitatively replicated one another (e.g., "Study A and Study B each showed gender differences in stress.") with different stress indicators as the basis for this general inference in each study. This level of analysis may, in some instances, ignore the fact that the two studies being compared included some variables in common (e.g., a depression scale) that showed a difference in one study, but not the other. When failures to replicate findings are noted, it is likely that they will be explained away as examples of specific effects affecting particular groups. The apparent specificity will lead to a call for studies of differences in the stress processes in the groups. The alternative that the results represent chance sampling variation or differences in sample size or other design artifacts ordinarily is not raised as an issue. The net result is that small, nominally significant differences that do not replicate will be mistaken for general stress effects when a focus on specific topics within the broad domain would be a more productive approach to the topics of interest in the research.¹⁹

Solutions to the preceding problems exist. Methods have been

developed to provide quantitative reviews of the evidence in a field, including tests for the influence of design and methods factors on the size of the estimates (Rosenthal, 1984; Hunter & Schmidt, 1990). The present semi-quantitative review of the evidence illustrates the need for and potential payoffs from more rigorous application of those integrative methods in this area.

Caveats

Balanced judgment of the state of the art in gender-stress research requires attention to factors that might bias the estimate of stress differences downward. Indicator reliability is one such factor. Many of the correlations reviewed were obtained using single item stress indicators or relatively short scales. Low reliability would lead to lower observed correlations (Nunnally & Bernstein, 1994). However, if the true population correlations were even as large as $r_{pb} = .30$, the indicator reliability would have to be $r_{xx} = .41$ to yield the observed median value of approximately $r_{pb} = .05$.²¹ The reliability of individual items might be this low.

The fact that correlations were not weighted in any way might bias the results. Two types of weighting might have been appropriate. First, no attempt was made to allow for differences in sample size. Second, each correlation reported in a study was treated as an independent data point. As a result, some studies contributed many observations to the analysis while others contributed a single association. The fact that some studies contributed only selected results to the analyses was even more important. This occurred when a study reported only the statistically significant findings.

The lack of weighting procedures may not substantially affect the inferences about gender differences. Comparing studies with relatively small samples ($n \leq 280$) to studies with larger samples, larger samples were nearly twice as likely (76.5% vs 40.7%) to produce gender differences in the $-.10 < r_{pb} < .10$ range. Also, a larger number of the gender difference estimates came from studies with large samples than studies with small samples (313 vs 123). Combining these two observations, weighting by sample size would have emphasized gender differences that were near zero.²²

Treating multiple differences from a single studies as separate data points may be a more significant problem. Studies with six or fewer variables tended to produce more substantial gender differences than studies with more than six variables. Treating each difference in the studies with more variables separately, therefore, gives more weight to some small differences than they would receive if they were averaged. For example, 10 small correlations between .00 and .10 would be replaced by a single small correlation. Averaging effect sizes within studies probably would have affected the overall distribution, but would not have affected the overall conclusion from the data. Gender differences tended to be larger in the studies with fewer measures. Averaging the findings within each study, therefore, would have meant that a large number of small correlations were combined to yield a smaller number of small correlations.

A crude estimate of the effects of averaging suggests that this procedure would have less impact than it appears at first. The distribution of results in the subset of studies with five or fewer measures was examined. This provided a closer approximation to the results than would be obtained if each study were reduced to a single estimate. Most (61.8%) gender differences in this data subset fell below the $r_{pb} = .10$ level. The estimated median correlation that was only trivially larger than that estimated for the full data set ($r_{pb} = .05$ vs $r_{pb} = .03$).²³

Considering these caveats and the evidence reviewed earlier, the biggest threats to the validity of the inference that gender differences in stress and strain are modest appear to be excessive generalization and unreliability. Excessive generalization refers to making statements about stress and strain as though they were homogenous general categories, despite the fact that some specific types of stress and strain may be consistently higher in women. Unreliability refers to the fact that many estimates of gender differences rely on individual items from stress questionnaires. This approach can be sensitive to even isolated differences in stress, but that potential may overlook any differences that do exist because the estimation of those differences depends on imprecise measurements. As research progresses, it will be important to replace broad coverage with more focused investigations of those aspects of stress that may truly differ between females and males. The narrower focus will permit more reliable measurement of the constructs.

Conclusions

Women and men generally report comparable levels of stress. This observation, initially made in an earlier review by Martocchio and O'Leary (1985), applies equally well to the present sample of subsequent findings on gender differences. If there is a mistaken perception that such differences are common and large, this mistake appears to be an example of common research practices that have been strongly criticized for some time (e.g., Meehl, 1978). Work on gender differences probably will proceed more effectively if the focus on the broad domains of stress and strain is replaced by efforts to understand specific sub-domains. Gender differences in negative affect, general distress, and work-family conflict are three areas that merit further consideration. A more complete, systematic meta-analysis of the available evidence on gender differences in stress could be a useful means of extending this observation, but such a review is beyond the scope of this study. Once completed, such a review would be expected to point to a limited number of specific aspects of stress and strain that should be investigated in more detail using more reliable assessments than are feasible in initial broad spectrum screening studies.

Literature Summarized

Anderson, E. A., & Leslie, L. A. (1991). Coping with employment and family stress: Employment arrangement and gender differences. *Sex Roles*, 24, 223-237.

Anshel, M. H., & Kaissidis, A. N. (1997). Coping style and situational appraisals as predictors of coping strategies following stressful events in sport as a function of gender and skill level. *British Journal*

of Psychology, 88, 263-276.

Bedeian, A. G., Burke, B. G., & Moffett, R. G. (1988). Outcomes of work-family conflict among married male and female professionals. *Journal of Management*, 14, 475-491.

Beutell, N. J., & O'Hare, M. M. (1987). Work-nonwork conflict among MBAs: sex differences in role stressors and life satisfaction. *Work & Stress*, 1, 35-41.

Brown, D. R., & Gary, L. E. (1987). Stressful life events, social support networks, and the physical and mental health of urban Black adults. *Journal of Human Stress*, 13, 165-174.

Brown, J., & Fielding, J. (1993). Qualitative differences in men and women police officers' experience of occupational stress. *Work & Stress*, 7, 327-340.

Burke, R. J., & Greenglass, E. R. (1989). Sex differences in psychological burnout in teachers. *Psychological Reports*, 65, 55-63.

Burke, R. J., & Richardsen, A. M. (1991). Sex differences in occupational stress and work satisfaction among physicians. *Stress Medicine*, 7, 79-86.

Caldwell, R. A., Pearson, J. L., & Chin, R. J., (1987). Stress-moderating effects: social support in the context of gender and locus of control. *Personality and Social Psychology Bulletin*, 13, 5-17.

Chaplain, R. P. (1995). Stress and job satisfaction: a study of English primary school teachers. *Educational Psychology*, 15, 473-489.

Clark, E. J., & Rieker, P. P. (1986). Gender differences in relationships and stress of medical and law students. *Journal of Medical Education*, 61, 32-40.

Conger, R. D., Lorenz, F. O., Elder, G. H., Simons, R. L., & Ge, X. (1993). Husband and wife differences in response to undesirable life events. *Journal of Health and Social Behavior*, 34, 71-88.

Cook, M., Young, A., Taylor, D., & Bedford, A. P. (1996). Personality correlates of psychological distress. *Personality and Individual Differences*, 20, 313-319.

Coverman, S. (1989). Role overload, role conflict, and stress: Addressing consequences of multiple role demands. *Social Forces*, 67, 965-982.

Cushway, D. (1992). Stress in clinical psychology trainees. *British Journal of Clinical Psychology*, 31, 169-179.

Fatkin, L. T. (1987). Daily reports of distress: Not for women only. *Work and Stress*, 1, 261-270.

Gadzella, B. M., Ginther, D. W., Tomcala, M., & W, B. G. (1990). Stress as perceived by professionals. *Psychological Reports*, 67, 979-983.

Gadzella, B. M., Ginther, D. W., Tomcala, M., & Bryant, G. W. (1991). Differences between men and women on stress producers and coping strategies. *Psychological Reports*, 69, 561-562.

Georgas, J., & Giadounaki, E. (1984). Psychosocial stress, symptoms, and anxiety of male and female teachers in Greece. *Journal of Human Stress*, 10, 191-197.

Gerdes, E. P. (1995). Women preparing for traditionally male professions: Physical and psychological symptoms associated with work and home stress. *Sex Roles*, 32, 787-807.

Gore, S., & Mangione, T. W. (1983). Social roles, sex roles and psychological distress: Additive and interactive models of sex differ-

ences. *Journal of Health and Social Behavior*, 24, 300-312.

Greenglass, E. R., & Burke, R. J. (1988). Work and family precursors of burnout in teachers: Sex differences. *Sex Roles*, 18, 215-229.

Greenglass, E. R. (1988). Type A behaviour and coping strategies in female and male supervisors. *Applied Psychology*, 37, 271-288.

Greenglass, E. R., Burke, R. J., & Ondrack, M. (1990). A gender-role perspective of coping and burnout. *Applied Psychology*, 39, 5-27.

Hackett, G., Betz, N. E., Casas, J. M., & Rocha-Singh, I. A. (1992). Gender, ethnicity, and social cognitive factors predicting the academic achievement of students in engineering. *Journal of Counseling Psychology*, 39, 527-538.

Hamilton, S., & Fagot, B. I. (1988). Chronic stress and coping styles: A comparison of male and female undergraduates. *Journal of Personality and Social Psychology*, 55, 819-823.

Hendrix, W. H., Spencer, B. A., & Gibson, G. S. (1994). Organizational and extraorganizational factors affecting stress, employee well-being, and absenteeism for males and females. *Journal of Business and Psychology*, 9, 103-128.

Heppner, P. P., Cook, S. W., Strozier, A. L., & Heppner, M. J. (1991). An investigation of coping styles and gender differences with farmers in career transition. *Journal of Counseling Psychology*, 38, 167-174.

Horwitz, A. V., White, H. R., & Howell-White, S. (1997). The use of multiple outcomes in stress research: A case study of gender differences in responses to marital dissolution. *Journal of Health and Social Behavior*, 37, 278-291.

Kaldenberg, D. O., Becker, B. W., & Zvonkovic, A. (1995). Work and commitment among young professionals: A study of male and female dentists. *Human Relations*, 48, 1355-1377.

Kandolin, I. (1993). Burnout of female and male nurses in shift-work. *Ergonomics*, 36, 141-147.

Kendler, K. S., Neal, M. C., Kessler, R., Heath, A. C., & Eaves, L. (1993). A twin study of recent life events and difficulties. *Archives of General Psychiatry*, 50, 789-796.

Lagace, R. R. (1988). Role-stress differences between salesmen and saleswomen: Effect on job satisfaction and performance. *Psychological Reports*, 62, 815-825.

Lester, D., Poser, I., & Leitner, L. A. (1986). Stress scores of workers by sex and age. *Psychological Reports*, 58, 110.

Levenson, H., Hirschfeld, M. L., Hirschfeld, A., & Dzubay, B. (1983). Recent life events and accidents: the role of sex differences. *Journal of Human Stress*, 9, 4-11.

Lewis, S. N. C., & Cooper, C. L. (1987). Stress in two-earner couples and stage in the life-cycle. *Journal of Occupational Psychology*, 60, 289-303.

Lichtenstein, P., & Pedersen, N. L. (1995). Social relationships, stressful life events, and self-reported physical health: Genetic and environmental influences. *Psychology and Health*, 10, 295-315.

McCleary, R., & Zucker, E. L. (1991). Higher trait- and state-anxiety in female law students than male law students. *Psychological Reports*, 68, 1075-1078.

McCreary, D. R., Wong, F. Y., Wiener, W., Carpenter, K. M., et al. (1996). The relationship between masculine gender role stress and

psychological adjustment: a question of construct validity? *Sex Roles*, 34, 507-516.

McIntosh, D. N., Keywell, J., Reifman, A., & Ellsworth, P. C. (1994). Stress and health in first-year law students: Women fare worse. *Journal of Applied Social Psychology*, 24, 1474-1499.

Milstein, M. M., Golaszewski, T. J., & Duquette, R. D. (1984). Organizationally based stress: What bothers teachers. *Journal of Educational Research*, 77, 293-297.

Murphy, J. M., Nadelson, C. C., & Notman, M. T. (1984). Factors influencing first-year medical students' perceptions of stress. *Journal of Human Stress*, 10, 165-173.

Newman, J. P. (1986). Gender, life strains, and depression. *Journal of Health and Social Behavior*, 27, 161-178.

Ott, E. M. (1989). Effects of the male-female ratio at work. *Psychology of Women Quarterly*, 13, 41-57.

Parkes, K. R. (1990). Coping, negative affectivity, and the work environment: Additive and interactive predictors of mental health. *Journal of Applied Psychology*, 75, 399-409.

Pilisuk, M., Boylan, R., & Acredolo, C. (1987). Social support, life stress, and subsequent medical care utilization. *Health Psychology*, 6, 273-288.

Pretty, G. M. H., McCarthy, M. E., & Catano, V. M. (1992). Psychological environments and burnout: Gender considerations within the corporation. *Journal of Organizational Behavior*, 13, 701-711.

Ptacek, J. T., Smith, R. E., & Zanas, J. (1992). Gender, appraisal, and coping: a longitudinal analysis. *Journal of Personality*, 60.

Ptacek, J. T., Smith, R. E., & Dodge, K. L. (1994). Gender differences in coping with stress: When stressor and appraisals do not differ. *Personality and Social Psychology Bulletin*, 20, 421-430.

Rayburn, C. A., Richmond, L. J., & Rogers, L. (1986). Men, women, and religion: Stress within leadership roles. *Journal of Clinical Psychology*, 42, 540-546.

Richard, G. V., & Krieschok, T. S. (1989). Occupational stress, strain, and coping in university faculty. *Journal of Vocational Behavior*, 34, 117-132.

Richardsen, A. M., & Burke, R. J. (1991). Occupational stress and job satisfaction among physicians: Sex differences. *Social Science and Medicine*, 33, 1179-1187.

Richman, J. (1984). Sex differences in social adjustment: effects of sex role socialization and role stress. *Journal of Nervous and Mental Disease*, 172, 539-545.

Richman, J. A., & Flaherty, J. A. (1990). Gender differences in medical student distress: Contributions of prior socialization and current role-related stress. *Social Sciences and Medicine*, 30, 777-787.

Romelsjo, A., Hasin, D., Hilton, M., Bostrom, G., Diderichsen, F., Haglund, B., Halqvist, J., Karlsson, G., & Svanstrom, L. (1992). The relationship between stressful working conditions and high alcohol consumption and severe alcohol problems in an urban general population. *British Journal of Addiction*, 87, 1173-1183.

Rosenfield, S. (1989). The effects of women's employment: Personal control and sex differences in mental health. *Journal of Health and Social Behavior*, 30, 77-91.

Rosin, H., & Korabik, K. (1995). Organizational experiences and propensity to leave: A multivariate investigation of men and women managers. *Journal of Vocational Behavior*, 46, 1-16.

Roxburgh, S. (1996). Gender differences in work and well-being: Effects of exposure and vulnerability. *Journal of Health and Social Behavior*, 37, 265-277.

Salgado de Snyder, V. N., Cervantes, R. C., & Padilla, A. M. (1990). Gender and ethnic differences in psychosocial stress and generalized distress among Hispanics. *Sex Roles*, 22, 441-453.

Savery, L. K. (1990). Men and women in the workplace: Evidence of occupational differences. *Leadership and Organization Development Journal*, 11(2), 13-16.

Sekaran, U. (1986). Significant differences in quality-of-life factors and their correlates: A function of differences in career orientations or gender? *Sex Roles*, 14, 261-279.

Sherman, A. C., & Walls, J. W. (1995). Gender differences in the relationship of moderator variables to stress and symptoms. *Psychology and Health*, 10, 321-331.

Simon, R. W. (1992). Parental role strains, salience of parental identity and gender differences in psychological distress. *Journal of Health and Social Behavior*, 33, 25-35.

Simon, F., & Corbett, C. (1996). Road traffic offending, stress, age, and accident history among male and female drivers. *Ergonomics*, 39, 757-780.

Sorenson, G., & al, e. (1985). Sex Differences in the relationship between work and health: the Minnesota Heart Survey. *Journal of Health and Social Behavior*, 26, 379-394.

Sowa, C. J., & Lustman, P. J. (1984). Gender differences in rating stressful events, depression, and depressive cognition. *Journal of Clinical Psychology*, 69, 615-622.

Spielberger, C. D., & Reheiser, E. C. (1994). The Job Stress Survey: Measuring gender differences in occupational stress. *Journal of Social Behavior and Personality*, 9, 199-218.

Swanson, V., Power, K., & Simpson, R. (1996). A comparison of stress and job satisfaction in female and male GPs and consultants. *Stress Medicine*, 12, 17-26.

Thoits, P. A. (1987). Gender and marital status differences in control and distress: Common stress versus unique stress explanations. *Journal of Health and Social Behavior*, 28, 7-22.

Toves, C., Schill, T., & Ramanaiah, N. (1981). Sex differences, internal-external control, and vulnerability to life stress. *Psychological Reports*, 49, 508.

Turner, R. J., & Avison, W. R. (1989). Gender and depression: Assessing exposure and vulnerability to life events in a chronically strained population. *Journal of Nervous and Mental Disease*, 177, 443-455.

Verbrugge, L. M. (1989). The twain meet: empirical explanations of sex differences in health and mortality. *Journal of Health and Social Behavior*, 30, 282-304.

Waelde, L. C., Silvern, L., & Hodges, W. F. (1994). Stressful life events: Moderators of the relationships of gender and gender roles to self-reported depression and suicidality among college students. *Sex Roles*, 30, 1-22.

Wheeler, R. J., & Frank, M. A. (1988). Identification of stress buffers. *Behavioral Medicine*, 14, 78-89.

Wu, X., & DeMaris, A. (1996). Gender and marital status differences in depression: The effects of chronic strains. *Sex Roles*, 34, 299-319.

Yeh, B. Y., Lester, D., & Tauber, D. L. (1986). Subjective stress and productivity in real estate sales people. *Psychological Reports*, 58, 981-982.

Zuckerman, D. M. (1989). Stress, self-esteem, and mental health: How does gender make a difference? *Sex Roles*, 20, 429-444.

APPENDIX B
Stress and Strain Items

The wording of the questions for stress and strain is given below. The response options for the items are described in the Methods section of the report. Items are given in the order in which they appeared in the questionnaire.

Strain Items

40. How do you feel about your:

- a. Job
- b. Personal life
- c. Health and physical condition
- d. Life as a whole
- e. Family
- f. (If married) Spouse
- g. (If you have children) Children

41. How many days during the past 7 days have you:

- a. Felt you just couldn't get going
- b. Felt sad
- c. Had trouble getting to sleep or staying asleep
- d. Felt that everything was an effort
- e. Felt lonely
- f. Felt you couldn't shake the blues
- g. Had trouble keeping your mind on what you were doing

Stress Items

42. Think about your whole life over the past 2 weeks. On the whole, how much stress do you think is in your life right now?

43. Of the stress that you experience, how much comes from problems or concerns with:

- a. Financial matters
- b. My personal health
- c. Personal or health matters of a family member
- d. Being aboard ship
- e. Crowded conditions aboard ship
- f. My personal safety aboard ship
- g. Maintaining personal hygiene aboard ship
- h. My lack of privacy aboard ship
- i. My inability to get enough exercise aboard ship
- j. The lack of recreational activities aboard ship
- k. My nutrition, the unavailability of desired foods aboard ship
- l. The person I work for (my immediate supervisor)
- m. The people I work with (my peers)
- n. The people who work for me (those I supervise)
- o. The way things are typically done aboard ship
- p. The people with whom I share living space aboard ship
- q. My ability to perform my duties
- r. My career and chances for promotion

Stress Items
(continued)

- s. Being able to stay in the Navy because of downsizing or force reductions
 - t. My relationship with my spouse or boyfriend/girlfriend
 - u. Breaking up with my spouse or boyfriend/girlfriend because of being aboard ship
 - v. My ability to communicate with my family and friends
 - w. Adapting to life after I return from this deployment
 - x. Feeling confined or trapped
 - y. My personal future and the meaning of my life
 - z. My use of alcohol
 - aa. Feeling isolated and excluded
 - bb. Feeling out of touch with the rest of the world
 - cc. My life as a whole
 - dd. (If you have children) My children because of being aboard ship
 - ee. (If you have children) Discipline of children
 - ff. (If you have children) Child-care arrangements
44. During the past 2 weeks, the stresses listed above have affected my:
- a. Personal life
 - b. Performance in my job
45. During the past 2 weeks, how well have you coped with these stresses?

APPENDIX C

Within-Type Moderator Analyses

Within-type moderator analyses tested the hypothesis that the size of gender differences varied among ships within a ship type. If not, all ships within a type can be accurately described by the average gender difference for their type. The alternative is that ship-specific gender differences would have to be computed to adequately understand shipboard stress and strain.

Within-type analyses were limited to four ship types, Fast Combat Support Ships (AOE), Ammunition Ships (AE), Destroyer Tenders (AD), and Submarine Tenders (AS). Only these four types satisfied two conditions required for meaningful analysis. First, at least three ships within the type were needed to reduce the risk of concluding that within-type effects were present merely because one representative ship from a generally homogenous type was compared to an outlier ship for that type. Second, all ships in the analysis had to meet minimum sample size requirements for females and males separately. The a priori objective was at least 20 in each group. This criterion was relaxed to a minimum of 18 in each group after examining the data. Relaxing the criterion added two ships to the set analyzed and increased the number of types analyzed from three to four.

Three ship types (AOE, AE, and AD) were similar within each type. The cumulative χ^2 was statistically nonsignificant for each of these types (AOE, $\chi^2 = 101.04$, 100 df, $p < .395$; AE, $\chi^2 = 100.41$, 100 df, $p < .447$; AD, $\chi^2 = 95.11$, 100 df, $p < .839$).

ASs demonstrated highly significant within-type variation in the size of gender differences in stress and strain ($\chi^2 = 475.40$, 300 df, $p < 5 \times 10^{-10}$). This result was primarily attributable to a single outlier. Roughly one third of the total χ^2 was produced by AS-5 ($\chi^2 = 164.65$, 50 df, $p < .001$, Table C-1). That ship produced large effects for QoL ($|d| = -.17$), depression ($|d| = .32$), and stress ($|d| = .14$). Aboard this ship, eight items exceeded the item significance criterion, but those effects appeared to be no more than the extreme values in a larger distribution.

However, even with AS-5 excluded, the differences between ships still were significant ($\chi^2 = 302.33$, 250 df, $p < .014$). This overall result was primarily the result of differences in stress ($\chi^2 = 237.36$, 180 df, $p < .003$). The differences in QoL ($\chi^2 = 29.48$, 35 df, $p < .732$) and depression ($\chi^2 = 35.49$, 35 df, $p < .446$) varied by no more than chance amounts.

AS-3 and AS-6 differed from average with regard to stress (AS-3, $\chi^2 = 59.73$, 36 df, $p < .008$; AS-6, $\chi^2 = 60.13$, 36 df, $p < .008$) even with AS-5 removed from the analysis. Neither ship deviated from the average AS difference for QoL or depression ($p \geq .074$). Thus, the primary differences for these two ships were in the stress domain.

AS-3 produced significant effects for only 3 of the 36 stress items. Negative effects were obtained for concern about children ($z = -3.07$) and child care arrangements ($z = -2.75$). A positive effect was obtained for people with whom the person shared living space aboard ship ($z = 3.21$). Removing these three items from consideration left a trivial

Table C-1. Summary of Moderator Analyses for Submarine Tenders

Ship	Quality of Life		Depression		Stress	
	\bar{d}	χ^2	\bar{d}	χ^2	\bar{d}	χ^2
1	.01	3.90	-.08	6.50	-.08	53.24*
2	.09	6.35	-.07	4.47	-.02	26.47
3	-.02	3.33	-.20	15.37*	-.06	58.30*
4	.00	5.05	-.05	3.19	-.03	25.17
5	-.17	16.13*	.32	59.95*	.14	88.55*
6	.05	3.28	.03	6.35	.04	53.03*
7	.12	8.38	-.04	2.64	-.05	25.72

Note. Ship numbers refer to the order of the ship in the data set. These numbers do not correspond in any way to the hull numbers assigned to the ships by the U.S. Navy. Bold typeface indicates a statistically significant effect with $|ES| \geq .10$.

deviation from the within-type trends ($\chi^2 = 32.42$, 33 df, $p < .496$). There was no consistent tendency for z-scores to be positive or negative (20 negative vs 16 positive). These facts indicate that AS-3 had moderator effects limited to a few stress indicators. The effects that were noted were inconsistent with respect to the direction of the difference.

The general tendency for women to report more stress and strain than men was amplified aboard AS-6. Amplification was evident in positive effects for 28 of 36 items, but none of the individual effects was large ($|z| \leq 2.34$). As a result, the small average positive effect ($|d| = .06$) was representative of this ship. Given this effect size, the statistical significance of the trends for this ship can be attributed to the moderately large sample size for this ship ($n = 146$).

The within-type analysis supported the view that results obtained for overall ship type adequately represent gender differences for nearly all ships (13 of 16; 81.3%). One of the remaining three ships deviated significantly from the overall type profile only because a small effect size was combined with a large sample size. In a second case, the apparent deviation from the overall type profile was limited to 3 of 50 items. Using ship type as the basis for inferring gender differences aboard individual ships, therefore, would yield a substantial, general error in only 1 of 19 cases.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)

2. REPORT DATE

November 1998

3. REPORT TYPE & DATE COVERED

Final Dec 1994 - Dec 1997

4. TITLE AND SUBTITLE Organizational Influences on Gender
Differences in Stress and Strain Aboard U.S. Navy Ships

5. FUNDING NUMBERS

Program Element:
Work Unit Number: Army
Reimbursable-6611

6. AUTHOR(S) Vickers, R.R., Jr., and Martin, J.A.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Health Research Center
P.O. Box 85122
San Diego, CA 92186-5122

8. PERFORMING ORGANIZATION

Report No. 98-33

9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)

Bureau of Medicine and Surgery
Code 26
2300 E Street NW
Washington DC 20372-5300

10. SPONSORING/MONITORING
AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The work was funded by the U.S. Army Medical Research and Materiel Command, Fort Detrick, Frederick, MD 21702-5012. Dr. Martin is at the School of Social Work, Bryn Mawr College, Bryn Mawr, PA

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words) Changes in policy have resulted in more women being assigned to U.S. Navy ships in recent years than has been true historically. This study tested the general hypothesis that women would find these assignments no more stressful than men do. This hypothesis, based on a review of stress studies in civilians, was supported. The review of studies of civilians suggested that work-family conflict would be an exception to the general trend, and women did report substantially higher concern about children and child care than did men. Gender differences in depressive symptoms were trivial even though studies of civilians suggested they would be large enough to be of interest. The difference may be that the women and men in this study were matched on sociodemographic attributes that could affect depression. Occupation (rating), ship type, and deployment experience were considered as organizational factors that might increase the size of the typical gender differences in stress and strain. None of these factors substantially influenced the size of gender differences in stress. Some specific instances of larger than average gender differences in stress and strain were noted, but these isolated trends require replication before they can be given much weight. On the whole, gender differences in stress were small and comparable to those in civilians.

14. SUBJECT TERMS psychological stress gender
psychological strain shipboard
military personnel military

15. NUMBER OF PAGES
46

16. PRICE CODE

17. SECURITY CLASSIFI-
CATION OF REPORT
Unclassified

18. SECURITY CLASSIFI-
CATION OF THIS
PAGE
Unclassified

19. SECURITY CLASSIFI-
CATION OF
ABSTRACT
Unclassified

20. LIMITATION OF
ABSTRACT
Unlimited